2015/2016

PRORAIL B.V.
Moreelsepark 3
3511 EP
Utrecht
Info@ProRail.nl

UNIVERSITY OF TWENTE
Drienerlolaan 5
7522 NB
Enschede
Info@Utwente.nl

NIELS.DEWIT
Scheidingsvliet 4
1759 LC
Callantsoog
Nielsdewit92@Gmail.com

DEVELOPMENT OF A RELIABLE PREDICTION METHOD FOR URGENT INFRA-FAILURE RECOVERY TIMES AT PRORAIL B.V.

A Report about the Prognosis of the Repair Time and an Accurate and Reliable Timeslot Determination of Urgent Failure Recoveries in the Railway Industry.

Development of a Reliable Prediction Method for Urgent Infra-Failure Recovery Times at ProRail B.V.

Niels de Wit

Coordinators University of Twente:

Ahmad Al Hanbali & Peter Schuur

Coordinators ProRail B.V.:

Paul Ammerlaan & Peter Franken

September 2015 – February 2016

Preface

You have before you my master-thesis about the subject: "Development of a Timeslot Length Prediction Method of Urgent Infra-Failures at ProRail B.V.": my final project as a student. The last couple of months of my study: Industrial Engineering & Management I spent at the wonderful train-industry world at ProRail in Utrecht. This offered a unique option to apply my theoretical knowledge in a practical situation in a world closely watched and judged by the entire country. I did not just learn a lot of this but I also met lots of amazing people willing to help me at all times and making me enjoy my stay at ProRail a lot!

I would like to thank my coordinators at ProRail: Peter Franken and Paul Ammerlaan for the amount of support and feedback I received on my research. The enthusiasm I was welcomed with in the team and the willingness to apply my research in practice really motivated me to bring this research to a successful end. A special thanks to Etienne Weijers, Saskia Wevers, Hans Kuijlen and my colleagues from Meldkamer Spoor for always helping me with lots of questions and helping me with my report. This helped me to take the report to a higher level. I also want to thank my coordinators of the University: Ahmad Al Hanbali and Peter Schuur. They left me free to execute my research making sure to point me in the right direction with good feedback whenever necessary. Thank you very much for this!

I hope you enjoy reading this thesis!

Niels de Wit

Utrecht, 4th of March 2016

List of Abbreviations

During this research multiple abbreviations are used. Some of these are frequently used at ProRail itself and, to prevent confusion, therefore not changed into English abbreviations. The English term of this is given afterwards.

ATB: "Automatische Treinbeïnvloeding" (Automatic train stopping device).

BO: Back Office.

DOT: "Dringende Onregelmatigheid met Tijdsafspraak" (Urgent Failure with a Time-appointment).

DVL: "Decentrale Verkeersleiding" (Decentralized Traffic Control).

FRM: Failure Recovery Moment.

FRT: Failure Recovery Time.

LVL: "Landelijke Verkeersleiding"(National Traffic Control).

MKS: "Meldkamer Spoor" (Emergency Room Railway).

MTTF: Mean Time To Failure.

MTTR: Mean Time To Repair.

NIC: "Niet In Controle"(Not Under Control).

NS: "Nederlandse Spoorwegen" (Dutch Railways).

NUO: "Niet Urgente Onregelmatigheid"(Non-Urgent Failure)

OBI: "Operationeel Besturingscentrum Infra" (Operational Control Centrum Infra).

PCA: "ProcesContractAannemer" (Contractor).

Stoco: "Storingscoördinator" (Failure Coordinator).

TAO: "Trein Aantastende Onregelmatigheid" (Train Affecting Irregularity).

TIS: "Trein Incident Scenario" (Train Incident Scenario).

TOBS: "Ten Onrechte Bezet Spoor"(Train Vacancy Detection Failure).

Trdl: "Treindienstleider" (Movement Inspector).

UO: "Urgente Onregelmatigheid"(Urgent Failure).

VL: "Verkeersleiding" (Traffic Control).

WBI: "Werkplek Beveiliging Instructies" (Workplace Safety Instructions).

List of Definitions

BO: The BO is responsible for the caretaking of calamities and incidents.

DOT: Generally classified at ProRail as a priority 5 failure. There is hindrance for the trains, security issues or damage but there is no direct access to the infra (thus a time-appointment is necessary).

District Control Office: The thirteen district control offices takes care of their train traffic in a specific region. In Dutch it is a "VL-post".

Diamond Crossing With Double Slips: A switch capable of transferring a train over another path, like a crossing (in Dutch: "Engels Wissel").

Failure form: A method of failure for a specific type of infra-equipment.

FRT: The time between the arrival of the mechanic at the rail track and the time the mechanic exits from the rail track after the repair action.

FRM: The moment when a failure recovery takes place.

Full repair: A type of repair permanently (till the next failure) restoring the infra-equipment to its full capacity.

Infra-equipment: A type of equipment from ProRail's infrastructure, for instance, a switch or a path (the tracks used by the train).

MKS: MKS is an experimental department active in the area of The Hague (September, 2015) aiming to create one department (MKS) capable of handling the failure notification intake.

MTTF: The average time between two failures.

MTTR: The average time it takes to repair a failure.

NIC: A term used to indicate that an object (usually a switch) is not under control and thus, does not respond. This is a common failure form.

NUO: Generally classified at ProRail as a priority 4 failure. There is no hindrance for the trains, security issues or damage and there is no direct access to the infra necessary.

OBI: Responsible for failure notifications requiring access to the infrastructure. They also take care of the energy supply of trains.

Path: The path (in Dutch: "spoor") is one railroad track used by a train.

Path withdrawal: During a path withdrawal ("spooronttrekking") the train traffic schedule is altered such that a certain path is no longer used in the schedule for a period. This enables a mechanic to execute a failure recovery in that path. Due to safety reasons a path withdrawal usually includes multiple paths at once. It can be activated by the VL.

Route Section:	A route section (Dutch: "Baanvak") is a part of the path between two stations or other relevant points for the train schedule.
TAO:	A failure which influences the train traffic. For instance: a switch which can't be controlled.
Temporary repair:	A, usually faster, type of repair which only repairs the failure temporarily. Thus, the infra-equipment can be fully used but only temporarily. A full repair is necessary within a specific amount of time. This type of repair is typically used when full capacity of the infra-equipment is necessary on the short term.
TIS:	A TIS is a standard scenario corresponding with certain characteristics of the failure, a first prognosis and actions which have to take place.
Trdl:	The Trdl ("treindienstleider" or movement inspector) works for traffic control and is responsible for the safety of the railway path in his own appointed area. He has direct contact with the train drivers about access to certain tracks and with mechanics about path withdrawals. The Dutch term: treindienstleider is used in this report.
Train Vacancy Detection Failure:	A train vacancy detection failure happens when a section in a path registers a train while there is none. This is a security system failure. In the remainder of the research the Dutch term: TOBS is used.
UO:	Generally classified at ProRail as a priority 1 or 2 failure. There is hindrance for the trains, security issues or damage which requires direct access to the infra.
WBI:	This is a description of the actions necessary for maintenance to a certain infra-equipment containing, for instance, the paths that have to be blocked and the signs necessary to be placed).

Management Summary

ProRail B.V. is the company responsible for the entire railway infrastructure in the Netherlands: one of the densest railway networks in the world. An important goal of the company is to offer safe and reliable railway paths to the railway undertakings (for instance: NS and Arriva). In order to reach this, ProRail has to take care, via the contractor, of the infrastructure failures which, directly or indirectly, result in hindrance for passengers. The right way to do this is thus of vital importance for reliable passenger transportation.

Motivation: When a failure occurs, traffic control has to decide when to repair this failure. A direct repair of the failure and thereby a faster recovery of the path, requires paths not being available during the day while a postponement of the failure recovery to the night leaves traffic control to deal with the failure for a longer time. Choosing the best option of these two, with the least amount of passengers hindrance, is now done based on experience. This frequently results in the wrong decision made. This research develops a first step towards a decision tool automatically making the optimal decision with the least amount of passengers hindrance per failure. This is done by predicting one of the most important factors influencing the passengers hindrance: the failure recovery time (FRT). This FRT is defined in this report as the time between the arrival of the mechanic at the failure location and the announcement of the sign: failure recovered. More specifically, the research answers the following research question:

> *How can the length of the timeslot necessary to repair urgent failures for two types of repair types, namely full and temporary, be determined and used to support the decision making of the failure recovery moment?*

Answering this question has, besides the better decision making when the failure recovery moment has to be determined, more advantages:

- More reliable failure recovery information can be given to the passengers.
- A more suitable and reliable length of the failure recovery interval given to the contractor is determined.

Methodology of research: In the current situation, the FRT is predicted based on relatively old historic data per failure form. This prediction comes in the form of an average. The accuracy and usability of this number are checked in this report. We quickly determine that the high variability of the FRT per failure form prevents an accurate prediction based on the average of the failure form only. For instance: it overestimates the duration of a level crossing failure in 72% of the cases. This is also true for 62% of the TOBS-failures. In order to decrease this variability, new variables need to be used to predict the FRT. Besides this, the prediction should come in the form of an interval, instead of a value in order to make it more useful and reliable. When, for instance, 80% of the level crossing failures are repaired in 50 minutes, the traffic control is capable of basing decisions on this.

After the diagnosis of the performance of the current prognosis the report identifies possible FRT prediction methods to improve this prediction. We determined a list of possible methods based on research of scientific papers. From this list, the methods most suitable to be tested are selected based on various criteria (for instance: accuracy, speed and simplicity). The selected methods which are applied on the database and are the most promising to result in a good

prediction model are: confidence intervals via the probability distribution, regression analysis, prediction by the expert and nearest neighbour.

Per previously mentioned method a prediction model is set up. We next determined their performance by applying each model on a new, unused, set of data. Based on this, a model combining the confidence intervals via probability distribution and the relevant regression variables shows the best performance.

So far, only known variables are applied, for example; the failure form and the contract type. We further improved the performance of the FRT prediction by identifying new, relevant, and more specific variables. This is done by comparing different district control offices. These offices show different performance (in the form of the FRT) when confronted with the same situation. We visited multiple district control offices and identified and quantified the variables responsible for this difference. This enables us to give a more accurate FRT prediction for a more specific situation.

Results: Relevant variables for the FRT identified, besides the failure form, are: movability around the path, contract type, age of the path, contractor responsible, recent passing of the grinding train and a sharp corner directly before the TOBS. During a failure recovery it is possible to, based on the values of these variables and the probability distribution, determine the intervals relevant to base decision making on: the 50%, 80% and 90% percentiles and the success probability of the 25-minute intervals (the largest gap between train traffic).

Applying this in practice, at the district control office The Hague, in the form of a developed tool enables us to receive the first feedback of practical application and apply this. This feedback shows the need for the extension of a FRT-tool towards a tool capable of selecting the most suitable moment to repair the failure. In order to reach this, the link between the FRT prediction and the passenger hindrance per hour by the failure should be made. The latter is currently being developed (by employees of ProRail and the company: CQM) while the former is accomplished by this report.

This research has identified the factors necessary to predict the FRT of an infra-failure. It shows the difference between the expected FRT with a different PCA operating in the area, the influence of the grinding train on a TOBS and multiple other factors (quantified in the report). Moreover, we have tested the practical application of a FRT tool at the traffic control and the first feedback is acquired. Based on this, the possibility to combine the FRT tool with the passenger hindrance method (developed by CQM) is identified. With this link, a mathematically optimal moment for the failure recovery is selected. The actual development actions required to reach this are illustrated in a roadmap.

Roadmap: How to develop the tool from the current FRT prediction method to a fully functioning failure recovery moment tool requires multiple future steps. The roadmap displaying the steps which should be followed is shown in Table MS.1. This map continues from the starting point: first failure recovery time tool is being tested at district control office The Hague and ends when the complete failure recovery moment tool is active in The Netherlands. One main coordinator for the tool is assumed. Based on this roadmap the development and implementation is done in a coordinated fashion.

Table MS.1: The Roadmap to Nationally Implement the Failure Recovery Moment Tool.

Roadmap- Failure Recovery Moment - Tool

Step	Who? / How?
1. Automatically fill in Values for the FRT-Variables.	**Who?** Contractor and district control office The Hague. **How?** An infrastructure-expert of the contractor and the district control office will go through the infra-objects one by one. In the span of one or two days they can determine the values of the infra-oject for every relevant variable. Implement this.
2. Receive and Use First Feedback.	**Who?** Traffic control of district control office The Hague and coordinator tool. **How?** With a feedback session the traffic control can inform the coordinator of the tool about the limitations of the tool. These are analysed and improved.
3. Extend to Multiple District Control Offices.	**Who?** District control offices, contractors and coordinator tool. **How?** At every district control office, the need and relevancy of the tool will be clarrified. After this, the values for the variables of the infra-objects have to be clarified again. Next, the tool can be introduced one office at a time.
4. Develop Passenger Hindrance Tool.	**Who?** CQM and project leader hindrence-tool **How?** The current influence tool has to be extended with the WBI-influence and the hourly amount of delayed trains/passengers.
5. Combine Passenger Hindrance with FRT-Prediction.	**Who?** CQM, coordinator tool and project leader hindrence-tool. **How?** The objectives of the tool have to be clarified first. CQM or ProRail experts can next combine the seperate tools to fulfill the objectives.
6. Implement Combination at one District Control Office.	**Who?** District control office The Hague and coordinator tool. **How?** With the help of clear courses and instructions the use and value of the tool can be introduced at the office. Afterwards, an expert has to be present in order to help traffic control with the tool and to anser questions.
7. Receive and Use First Feedback.	**Who?** District control office The Hague, CQM and coordinator tool. **How?** The first feedback is received from the control office. CQM and the cooridinator of the tool together verify, analyse and improve the tool based on this.
8. Extend the tool to Multiple District Control Offices	**Who?** District control offices and coordinator tool. **How?** The need and relevany of the tool have to be explained at every district control office. Again, clear courses and instructions are used to introduce the tool one office at a time to the rest of The netherlands. Attention should be paid to feedback and questions about the tool.

Table of Content

1. Research Proposal

"Ladies and gentlemen, due to a switch failure, we are unable to continue our journey to Amsterdam Central. We will therefore return the train to Hilversum". When you arrive back at the station in Hilversum you are welcomed by the big blue signs with the text: "+30 minutes" for every train in the direction of Amsterdam. But after half an hour the sign displays "+60 minutes" and this keeps increasing every five minutes. A situation many a man in The Netherlands will recognize. This research aims to improve this unpleasant situation.

This research is written as a graduation project for the master: Industrial Engineering and Management with the specialisation: Production and Logistics Management. During six months the failure recovery times in the railroad industry are analysed with the cooperation between ProRail and the University of Twente. The project started at September 2015 and, with a time limit of six months, is planned to be finished in February 2016. In this chapter, the problem which is solved during this report is introduced and the relevance of solving it explained. This chapter also shows the methodology which forms the foundation of solving the problem and explains the structure which is followed throughout the rest of the report.

It is important to note that this essay is written entirely in English, however, due to the relatively national-oriented company, some of the Dutch terms and abbreviations are used to prevent confusion. The reader can find the explanation of these terms and abbreviations directly after the title page.

1.1. Company Description

ProRail is an independent business and the economic owner of the railway infrastructure in The Netherlands. It is responsible for the railways maintenance, security, extension and control. With 7,033 km railway, 2,731 railway crossings and 405 stations in The Netherlands ProRail takes care of one of densest railway infrastructures in the world, leading to an immense organization consisting of 4,000 employees (2014).

The origins of ProRail lie in a reorganization of the "Nederlandse Spoorwegen" (NS) in 1995 in which NS Railinfratrust B.V. was founded. This reorganization was done based on EU regulations which obliged splitting infrastructure and exploitations. Railinfratrust was formed and removed the more government-oriented activities from the commercial company NS. In 2002 Railinfratrust was removed from the NS holding and the company is called ProRail since 2003. Since 2009 the company consists out of three main units: projects, operations and transport & scheduling.

The main goal of ProRail is to deliver reliable and safe railway paths to railway undertakings ("vervoerders"). Since ProRail is responsible for the transportation of hundreds of thousands of people per day, failures on the path ("spoor") will directly result in lots of delayed persons which results in political pressure, extra costs and damage to the image of the train industry in general. To prevent this, ProRail is constantly improving its failure recovery system in order to be able to reduce the amount of failures that have impact on train schedules and punctuality.

The operational maintenance of the path is not done by ProRail itself. For the execution of maintenance activities, the company makes use of contractors like BAM Rail, Strukton Rail and Volker Rail. Prorail has a contract with these companies giving them responsibility for the execution of the maintenance activities on a certain part of the railway paths in the Netherlands. The contractors are rewarded either performance- or activity-based. The transportation of passengers is done by railway undertakings like NS and Arriva, making use of the railway paths offered by ProRail. These railway undertakings pay a fee to ProRail for using the infrastructure. Thus, ProRail can be seen as the facilitator between the railway undertakings and the contractors (see Figure 1.1).

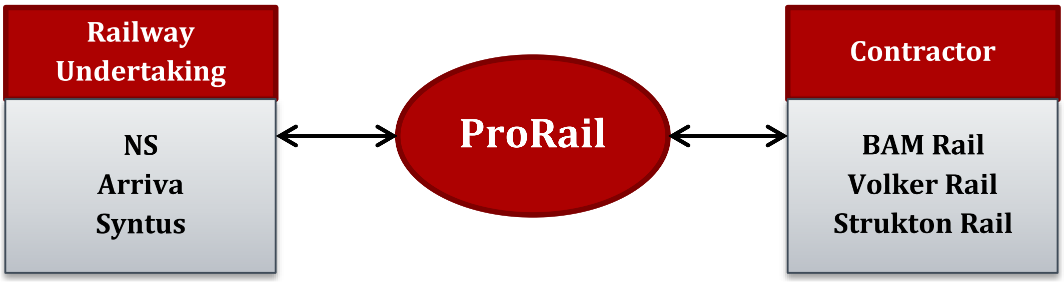


Figure 1.1: The Main Parties ProRail deals with.

1.2. Problem Description

In 2015 Prorail is running an experiment in The Hague called: "Meldkamer Spoor (MKS)". During this experiment, ProRail tries to create one department (MKS) which is responsible for and capable of handling the intake of the failure notification. Previously, this failure notification was first called in by the traffic control department to the "back office" (BO), responsible for handling incidents and calamities, while the intake of a failure which requires access to the infrastructure was done by calling to the "Operationeel Besturingscentrum Infra" (OBI) , responsible for the infrastructure. The OBI then calls to the representative of the contractor, the Stoco, to actually arrange the failure recovery. The MKS thus combines the functions of the BO and OBI such that only one call by the traffic control department has to be done. By reducing the amount of necessary contacts, the process becomes faster and less error-prone. However, during the planning process of the MKS, ProRail discovered that there are multiple obstructions that pose a threat for the expansion of MKS to a national level. These obstructions already existed for some time but were, so far, not noticed or not deemed important. An important obstruction is the planning when to do an urgent failure recovery action.

There are multiple types of failures recoveries ProRail is executing. These can be placed into three categories:

- UO (urgent failure or "Urgente Onregelmatigheid"), generally classified at ProRail as a priority 1 or 2 failure.
- DOT (urgent failure with an appointment or "Dringende Onregelmatigheid met Tijdsafspraak"), generally classified at ProRail as a priority 5 failure.
- NUO (non-urgent failure or "Niet Urgente Onregelmatigheid"), generally classified at ProRail as a priority 4 or 9 failure.

UO and DOT recoveries are for failures which directly result in hindrance for the train traffic and thus for passengers and should be repaired as fast as possible. These failures are always corrective maintenance[1]. For these recoveries, timing is extremely important and the failure should be handled as quickly as possible. NUO recoveries are based on failures which do not

[1] Corrective maintenance is "the unscheduled maintenance (or repair) to return an item/equipment to a defined state and carried out because maintenance persons or users perceived deficiencies or failures" (Dhillon, 2002).

result in hindrance for passengers and can therefore be done during a more suitable time in the coming days/weeks or be deemed not necessary at all. These failure recoveries can be either corrective or preventive maintenance[2]. NUOs are not problematic and are not be treated during this research.

The repair action of an urgent failure should be done as quickly as possible. However, for logistical reasons, direct repair is sometimes not possible or advisable. An example of this is a failure half an hour before the start of rush-hour. If the repair action is done as quickly as possible, multiple paths (due to safety reasons not only the path with the failure will be jammed, but also the paths around it) can't be used during rush hour. This leads to huge hindrance for passengers. The repair action is therefore often transferred to a more suitable time.

A failure during or just before rush-hour is problematic but, decision-wise, easy: don't repair it now because it won't be finished before rush-hour. However, making a decision for a failure recovery moment three or four hours before rush hour is more difficult: either repair it now and have a fully functioning system during the rush hour or repair it later but run the coming hours with an only partially functioning system. Thus: take the pain now or suffer for a while. It is also possible to use a hybrid solution: the mechanic only executes a temporary failure recovery. This is a faster failure recovery method but only solves the problem for a short amount of time. This problem is called: failure recovery moment (FRM) problem throughout this report.

Developing a tool to support this logistical decision-making process is the end-goal of this research subject. However, first, the most important input factor of this process should be researched: the necessary length of the timeslot for a failure recovery action of a contractor. This problem is referred to as: the failure recovery time (FRT) problem and is tackled in this report.

When a failure recovery with a time-appointment (DOT) will take place the contractor gets a timeslot in which he has access to the path and can fully repair the failure. During this timeslot, trains do not have access to the path with the failure and the paths around it. In the case of the experiment of The Hague the timeslot can be the entire night without hindering any passengers since there are barely any trains riding during the night in The Hague. This convenient option is not possible for the rest of The Netherlands. Here, a careful consideration about the length of the timeslot has to be done. A timeslot which is too long leads to a loss of capacity while a timeslot which is too short leads to a disturbance of the planning when not sufficient. All of this resulting in unnecessary extra delays for the passengers.

In order to keep these delays to a minimum the timeslot has to be as short as possible while still being able to guarantee that the contractor is capable of repairing the failure within the timeslot. This not only helps to shorten the closing of the paths but also forms the basis for the decision-making support tool for the moment of failure recovery stated previously. Besides this, it is also a good indication for the length of the failure repair for an UO (urgent failure) and can therefore be used to inform passenger more accurately about the length of delays.

To be capable of reliably determining the appropriate length of the timeslot it is necessary to predict the length of the repair action per failure form (the term used at ProRail) for both

---

[2] "Preventive maintenance actions are carried out on a planned, periodic, and specific schedule to keep an item/equipment in stated working condition through the process of checking and reconditioning" (Dhillon, 2002).

temporary and full repair. The failure form is the initial, quite generally formulated, failure mode of an infra-equipment type (a type of equipment from ProRail's infrastructure, for instance, a switch or a railway path). An example of a failure form is: switch not under control (a non-responding switch). There are however, variables or values in this determination that are not known to ProRail. An example of this is the time a mechanic has to wait after arrival before he can access the track. Improving these and making them suitable to be used in the failure recovery process is a major step in this research.

Thus, this research supports the development of a tool for the decision making process of the traffic control for the failure recovery moment: the FRM problem. In order to be able to develop this, however, a major obstruction has to be tackled: the determination of the length of the timeslot necessary for the failure recovery action: the FRT problem. If there is sufficient time, recommendations for the FRM problem are made as well.

By conducting multiple interviews with relevant parties the problem tree of the research is determined. The problem tree is shown on the next page in Figure 1.2. In order to keep the tree manageable, only relevant sub-problems for the research are shown. These sub-problems are thus the relevant problems for solving the FRM problem via the FRT problem. Some of the cause and effect relationships may not be that clear just from the picture. Therefore, each of these gets a letter in Figure 2 and is explained, based on this letter, in Table 1.1.

Table 1.1: Explanation of the Unclear Cause-Effect Relationships.

Cause-effect relation	Explanation
A	The FRT is largely increased when a mechanic has to wait with the repair action because the traffic control did not close the train tracks on time.
B	The FRT is largely increased when a mechanic has to go back to get other tools
C	Different failure-types are categorized into the same category. For instance: a torn switch and a switch not under control are both categorized under: switch-failure while individual FRT's.
D	Different types of failure recovery can result in different times.
E	The FRT determination method does not research the possible causes.
F	The contractor may have more insight into the causes but he does not share this with ProRail due to a lack of data exchange.
G	ProRail does not exchange enough information about the failure and the access times leading to miscommunications.

Explicating the highest-level problem (the main-problem) to the lowest-level problem (the core-problems) results in six remaining core-problems. These are numbered in Figure 1.2. From these six, only one is not relevant for this research: core-problem four. Increasing the amount of data collection at ProRail will, relatively, cost a lot while it does not pose large benefits on the short term since the data which is already collected is not used and/or collected effectively. The other core-problems are all being treated during this research. First, core-problems: one, two, three and five are treated in order to determine the suitable lengths of the timeslots. Combining these with a solution for core-problem six creates a complete picture of the possibilities and results in the best way to effectively and efficiently solve the main-problem.

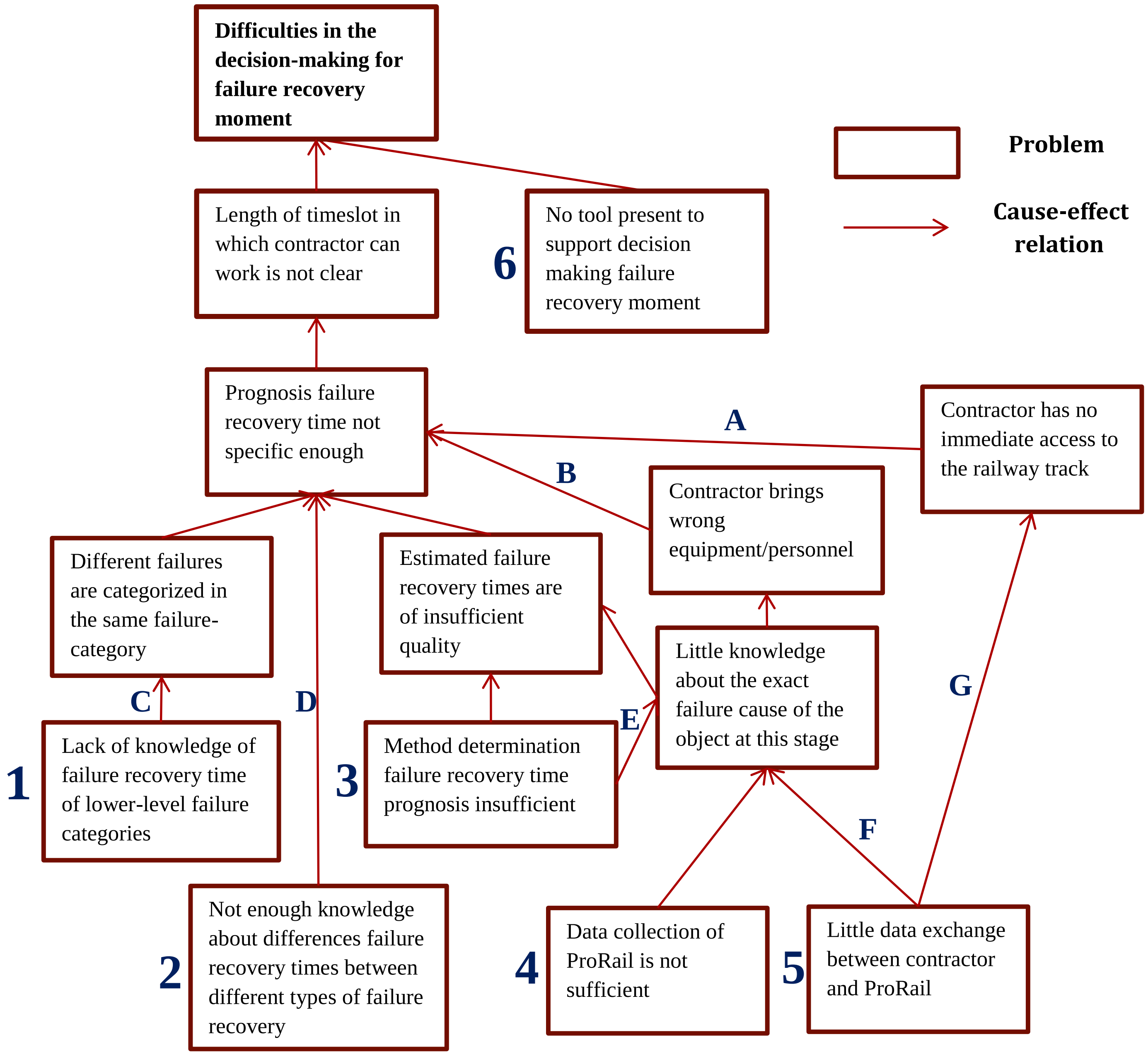


Figure 1.2: The Problem Tree of the Research

1.3. Research Goal

From the problem description the following goal can be formulated:

> *Developing a decision support tool for the failure recovery moment based on a reliable first prognosis of the necessary length of the timeslot for the system recovery time for an urgent failure.*

Reliable first prognosis: The first prognosis is done directly after the failure notification is received. Comparing the final finishing time with the time predicted by the prognosis will determine whether or not the prognosis is reliable. To draw this conclusion, a base level of correct prognosis is necessary, meaning: at least X% of the prognosis are correct (the value of X is determined in a later stage of the research) to call it reliable. Furthermore, boundaries to call an individual prognosis correct or incorrect should be clear.

An important remark on this subject is the observation that the research is not just descriptive and focused on the prediction, but is also explanatory. This follows from the fact that the research aims to make this prediction more reliable by applying methods to reduce the variability in the FRT variability (for instance, by cause prediction).

System recovery time: The MTTR is the scientific term which corresponds closest to the system recovery time (the term used by ProRail). MTTR is defined as: "a basis measure of the maintainability of repairable items, it represents the average time required to repair a failed component or device" (BusinessDictionary, 2015). To be able to apply this definition on ProRail it is important to see from which point in time the MTTR should be measured and till when.

Urgent failure: Determining when to classify a failure as urgent and when not to is an important step for the research. According to the OBI, a failure which requires a system recovery to prevent damage, injury, delay or risk for simultaneous appearances of other failures is counted as urgent.

1.4. Scope of the Research

This research focuses upon support for the decision making of the traffic control for the FRM of the urgent failures. These type of failures are the source of the motivation for this research due to the relevancy of the decision when a time appointment has to be made on short notice. For all failure forms relevant a length in time is given, which can be used to base the timeslot for DOT (urgent failure with time-appointment/ priority 5 failures) failures on. The prediction made for the FRT can also directly be used for the prognosis for the final recovery time of UOs (urgent failures/priority 2 failures) and the FRM problem. It is only done for the failures where there is damage to the infra-equipment and thus a repair is necessary. Thus, a collision with a person is not included since this does not damage the infra-equipment.

Failures on the catenary ("bovenleiding") are not considered in this research. According to experts, these failures always take a lot of time to solve. When the catenary fails it is therefore always repaired during the night (or, when there is no other train traffic in the area possible anymore, it is done directly). The FRM (failure recovery moment)-problem therefore does not apply on these failure forms.

The focus of the research is upon the bulk of the failure forms. Thus, if the amount of data about certain failure forms proves insufficiently low (a low amount of annual failures for instance), this failure form is not included in the decision support tool. With this action, the largest part of the failures is taken care of and only a small fraction is skipped. The practical test is only done for one failure form. Excluding other type's results in a major time advantage while the test is still capable of tackling the most important practical problems.

The total time in which the train traffic is jammed during a failure is longer than just the FRT. However, this research only focuses upon the determination of the FRT. The remaining time is either beyond the direct reach of ProRail (the travel time of the mechanic to the location of the failure) or worked upon during other projects (project MKS for instance). If these times are necessary for further calculations, assumptions are made or constants are taken.

An important remark for the scope of the research is the focus on the transport of passengers instead of goods. The transport of goods in not considered in this research since they complicate the calculations a lot while they don't play a big role in the results since delays are not that problematic (compared to passenger trains).

1.5. Methodology

The research which is conducted is largely quantitative and contains both inductive (theory building) and deductive (theory testing) elements. A methodology suited for this is the "Algemene Bedrijfskundige Aanpak"(ABP) of Heerkens & van Winden (2012). This methodology distinguishes itself from other methodologies in its ability to treat multidisciplinary problems and the context of the company while maintaining its simplicity and flexibility. Since the problem affects multiple parties and departments a methodology which is capable of treating multidisciplinary problems while keeping it within the context of the company is very important. The flexibility and simplicity make it possible to create structural boundaries for a difficult problem without making it unmanageable.

Another important advantage of the methodology is the specialization of the theory in both knowledge- as handling-problems[3]. Since this investigation contains both the acquisition of knowledge as the application of it this is an advantage which matters.

The ABP-methodology makes use of the following seven steps:

1. Problem identification;
2. Formulation of the problem approach;
3. Problem analysis;
4. Formulation of alternatives ;
5. Decision;
6. Implementation;
7. Evaluation (Heerkens & van Winden, 2012).

These steps can be taken in both directions. Thus, it is possible to return to a previous step if this is deemed necessary. The ABP-steps form the basis of the structure which is followed during this research. The link between the structure of this research and the ABP is shown in Figure 1.3. As can be seen from the figure, the first two steps of the ABP are already done during this chapter.

[3] A knowledge-problem investigates an aspect of an environment, thus acquiring knowledge. A handling-problem actually changes this aspect (Heerkens & van Winden, 2012).

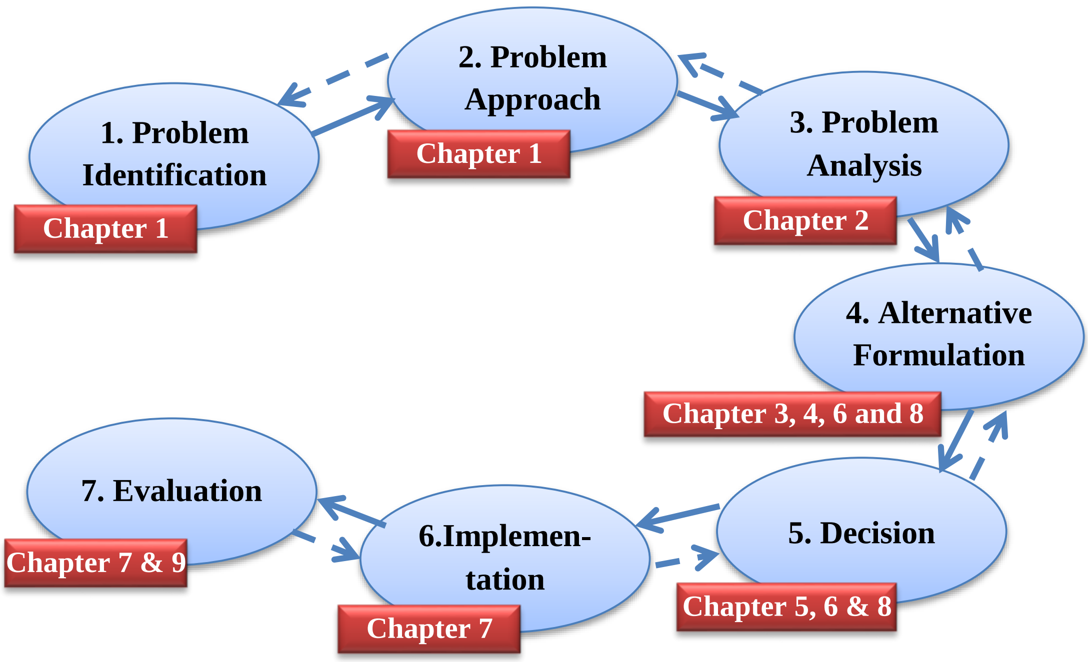


Figure 1.3: The Methodology linked to the Structure of the Research.

1.6. Research Questions

The research goal leads to the following main question which is answered during this research:

> *How can the length of the timeslot necessary to repair urgent failures for two types of repair types, namely full and temporary, be determined and used to support the decision making of the failure recovery moment?*

The main question is answered step by step by multiple research questions. The first of these research questions focus upon the FRT problem while the latter focus on the FRM problem.

1. **What is the current method to determine the necessary length of a timeslot and how does this method perform?** *(Step 3 of the ABP)*
    - 1.1. Which steps are taken during a failure recovery and by which parties?
    - 1.2. What is changed in the failure recovery steps in the experiment "Meldkamer Spoor"?
    - 1.3. How is the initial prognosis for a failure recovery duration done?
    - 1.4. How does this initial prognosis perform?

This research question identifies the current situation. This starting point indicates the current performance of the system and can therefore be used to look back on during the end-analysis in later research questions. Besides this, it identifies the data which is available and can be used in later analysis.

Answering this research question is done by making interviews with the relevant parties and looking into already written manuals and reports. The more quantitative questions can be answered by analysing the times of previous repair actions in the SAP-system (the most important system which is used to save the operational data). Data mining techniques (for instance: data cleaning and clustering) are used for the analysis of the database.

2. **What is a suitable method for the timeslot-prediction for a failure recovery per failure form?** *(step 4 of the ABP)*
    2.1. Which individual steps form the failure recovery time at ProRail?
    2.2. Which methods are suitable to determine the length of a timeslot for a failure form?
    2.3. Which method(s) is/are suitable for the situation at ProRail?

Answering this research question results in a number of methods which can be used to determine the length of a timeslot. This is compared with the situation at ProRail and the most suitable methods are selected. This question also identifies the individual steps within the failure recovery which influences the total time spend on the failure recovery. This research question thus answers core-problem three (see Figure 1.2).

The determination of the timeslot-prediction methods is done based on a literature study while the individual steps can be determined by conducting interviews with the relevant parties and looking into the database.

3. **How do the timeslot-prediction methods for a failure recovery perform when applied at ProRail?** *(step 4 of the ABP)*
    3.1. Based on the previously determined methods, what is the length of the timeslot of a failure form for each method?
    3.2. How should each method be represented in practice according to the stakeholders?

Answering this research question results in the lengths of the timeslots based on the methods determined in the previous chapter. Thus, this research questions examines core-problems: one and four (see Figure 1.2). A first test about the representation of the results in a practical situation is also done.

Applying the methods can be done by following the instruction from literature about it and combining this with data from, at least, the SAP-system. While a first draft of the results can be taken to the VL (traffic control) in order for them to form an opinion about the interface.

4. **What is the best failure recovery time prediction method?***(step 5 of the ABP)*
    4.1. How do these timeslot-prediction methods perform?
        4.1.1. What percentage of the first prognosis overestimates the length of the necessary timeslot?
        4.1.2. What percentage of the first prognosis underestimates the length of the necessary timeslot?
        4.1.3. What percentage of the time is the railway closed while this is not necessary?
    4.2. What is the improvement compared with the current situation?
    4.3. What is the best performing method?

This chapter tests the results of the methods applied before. New data is plugged into the methods and the results are compared with the predicted values. The best method can be selected based on its performance.

5. **In practice, how does the timeslot determination method perform?** *(step 6&7 of the ABP)*
    5.1. When given new failures, how often is the prognosis correct?
    5.2. How can the timeslot determination be further improved?

This research question contains the practical application of the timeslot determination made earlier. New failures are tested based on the new method and the prediction is checked for correctness. Thus now the method is applied while only the input is known instead of both the in- and out-put. This result is evaluated and further future improvements are given based on the problems encountered during the practical application.

This question is answered by looking at failures forms which come in and predicting, based on the model, what the FRT will be. When the true FRT turns out to differ significantly of the prediction, an analysis is done to determine why this was the case and what improvements have to be made to the timeslot-prediction model. The entire process chain of making the prognosis is followed and all the problems are thus encountered and, possibly, solved.

6. **For the same failure forms, what is the difference between the system recovery time of a full and temporary failure recovery?** *(step 4&5 of the ABP)*
    6.1. What is the difference between a full and temporary failure recovery?
    6.2. What is the motivation to execute a temporary failure recovery instead of a full failure recovery?
    6.3. What is the timeslot for a temporary failure recovery per failure form?
    6.4. How does this compare to the timeslot for a full failure recovery for the same failure form?

During this research question the timeslot determination is taken a step deeper: to different failure recovery types. By comparing the different timeslots of full and temporary failure recoveries of the same failure form, the decision-making for the failure recovery type is supported. This research question thus examines core-problem two (see Figure 1.2). This question is addressed in the appendix (Appendix I) since it is a sub-problem: the focus in this report is on the full failure recovery.

The difference between temporary and full recovery can be identified by analysing the SAP-data. The consequences of the differences are determined by presenting the results to the relevant parties.

7. **How to constantly improve the decision process of the failure recovery moment?** *(step 7 of the ABP)*
    7.1. What is the best method to improve a decision making process continuously?
    7.2. How to apply this method on the failure recovery moment selection process?

When the decision support tool is developed, it is important for traffic control to know how to use this tool and be evaluated on their use of it. By continuously improving this, the best decision can be made. This improvement process is set up by looking at the possibilities for this tool at ProRail. This is applied at ProRail next. Literature about decision making improvement tools can be used for the first sub-question while the latter is answered by taking interviews with the traffic control managers. The seventh research question is answered in appendix J. The question is postponed to the appendix since the focus of this report is on the prediction of the FRT, the extension to the failure recovery moment determination is a side-project.

1.7. Planning

The research questions are answered throughout the coming months. To set up a suitable planning, boundaries, in the form of number of days, are determined per sub-question. A detailed overview of this is given in Appendix A. This planning only shows the number of days necessary to anser the research question for the first time. Thus, it does not include the feedback time necessary to improve the ansering. This is not included since it is impossible to reliably determine this time. The time for free days and holidays is also excluded. Both of these points may probaly result in an extra month of work. This places the expected finishing time of the research at the end of February.

2. Performance of the Current Method

In this chapter the method which is currently used by ProRail to determine the necessary length of a timeslot is examined. The steps which are made in this process and the parties which play a role in each step are explained. The performance of the prognosis is also determined and is later on in the report used as a benchmark. An explanation of the MKS project is given to show the differences between the two situations and the effect this has on the use of the prognosis. These subjects are treated based on interviews with relevant parties in the failure recovery process and by use of the SAP-database.

2.1. Failure Recovery System – Current Situation

The process of failure recovery ProRail is currently using (thus, without MKS) consists of multiple parties and information transfers. An overview of the information transfers between the main parties in this process is given in Figure 2.1. As can be seen from the figure, it depends on the failure whether all the steps take place and which parties are informed. The way to read the decision tree and the parties is explained below the figure. This is followed by an explanation of the steps which take place during the failure recovery based on the communication processes indicated by the numbers in Figure 2.1

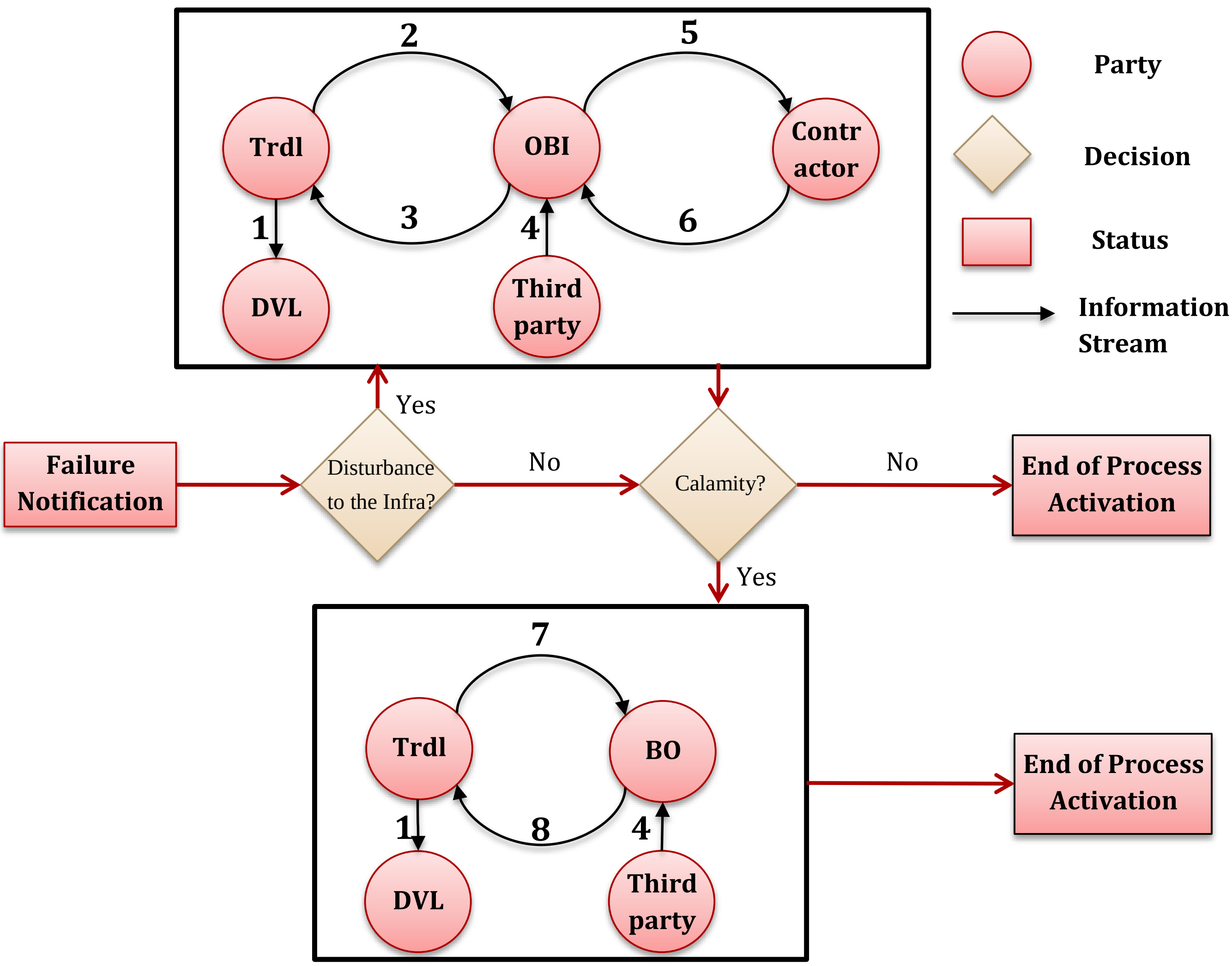


Figure 2.1: The Failure Recovery Process ProRail uses in the Original Situation

The numbers roughly indicate the order in which the activities usually take place. However, this order is influenced by the type of failure and method of failure notification, thus should only be used as an indication. The decision tree indicates which processes, indicated by the boxes, are activated. These boxes indicate continuous processes. These processes, and thereby: communication steps, keep going till the failure is solved. Thus, the end of the process is not influenced by the place in the decision tree of Figure 2.1 you are. In the case of a disturbance of the infra and a calamity, the OBI and BO processes are therefore activated at the same time.

An important decision which is made in the decision tree is whether you are dealing with a calamity or not. This is a difficult definition even within the company ProRail, therefore, another decision tree should be followed in order to determine this. This one is given in Appendix B and results in the classification of the failure as a calamity or a regular disturbance.

<u>Parties:</u>

- *Trdl (Treindientsleider or movement inspector):* The Trdl plays a central role in the transportation of passengers and goods over the railway paths. This person is responsible for the safety of the railway paths in his own appointed area and has direct contact with the railway undertakings. The failure notifications are received by the Trdl via his own control equipment or via the train drivers. The Trdl is part of the "verkeersleiding" (VL) or traffic control who are responsible for the logistics of the railway paths. The VL is enabled to actually stop the driving of trains.

-*DVL* (*"decentrale" or decentralized VL*): Where the trdl is the first level of the VL, the DVL is the second level. When a failure in the area of a Trdl influences multiple areas he contacts the DVL, who takes care of one of the thirteen different regions in The Netherlands and oversees the influences of different failures on the train traffic. When the failure impacts multiple regions, the DVL contacts the third level: LVL ("landelijk" or nationwide VL). Practically speaking, a failure almost always has an impact on different areas, making the DVL an important player in the failure recovery process.

- *BO (Backoffice):* The BO is responsible for the caretaking of calamities and incidents. They contact the institutions necessary to take care of these (for example: the police and mortician). Furthermore, they make sure that the information distribution about the development of the incident to the right parties is taken care of.

- *OBI (Operationeel Besturingscentrum Infra):* The OBI takes care of the more technical side of the infrastructure. Via various channels the OBI receives its failure notifications which is used to lay contact with the corresponding contractor of the infra-equipment. The OBI is responsible for the efficient and effective development of the failure recovery.

-*Third party:* This party plays a minor role in the failure recovery process. It is possible that a third party (a civilian or inspector for instance) notices a failure. This party can directly contact the OBI in this case to report the failure. This only happens occasionally.

- *Contractor:* The contractor is directly responsible for the maintenance and recovery of the infra-equipment in its contract area. The main contractors ProRail deals with are: BAM Rail, VolkerRail, Strukton rail and ASSET Rail. The person who keeps contact with ProRail and is responsible as coordinator within the organisation of the contractor is the "storingscoördinator" (Stoco) or failure coordinator.

<u>Communication steps:</u>

-*Process step 1:* At this point in time the Trdl has just received a failure notification. He usually receives this notification via his own operating-system and/or via the machinist of a train. The Trdl informs the DVL about the failure. The Trdl explains the location, the failure form and the expected amount of FRT. This enables the DVL to analyse the effects of the failure and possibly inform other parties. The DVL gives feedback about the impact of the failure which is important input for the Trdl to determine the relevancy of the failure but also the failure recovery moment.

-*Process step 2:* The department the Trdl contacts here is the OBI. During this conversation the Trdl explains the location and gives a description of the failure form. The Trdl and the OBI also communicate about whether they are dealing with an urgent failure or not based upon the type of failure and the logistical possibilities and difficulties when dealing with this failure.

-*Process step 3:* The OBI gives feedback to the Trdl concerning the situation, explaining the priority level the OBI thinks is necessary for this failure. During the failure recovery the OBI keeps the Trdl aware of the progress of the failure recovery.

-*Process step 4:* Occasionally a third party notices a failure of an infra-equipment. In this case, the third party takes over the initial role of the Trdl by calling in the suitable party. In the case of a civilian, this is usually the OBI (the phone-number of the OBI is displayed near some of the infra-equipment). However, the third party being a train driver can result in a call to the BO instead since the train driver recognises the failure as a calamity only. The OBI or BO report this to the Trdl afterwards.

-*Process step 5:* During this step the OBI makes contact with the contractor responsible for the failed infra-equipment, to be precise: with the Stoco. This happens after the OBI received the failure information from the Trdl or third party. The situation and the priority given to the situation are explained. If the priority indicates that a timeslot for the failure should be arranged the OBI gives the contractor this timeslot. From here on, the OBI mainly monitors the failure recovery and occasionally steers the recovery process. If they receive more information about the failure recovery from the contractor, the OBI passes this on to the Trdl.

-*Process step 6:* The communication from the contractor towards the OBI mainly comprises updates concerning the failure recovery. The first thing the OBI receives is the notification that the mechanic has arrived at the failure location. After the mechanic has done his diagnosis, the mechanic gives the OBI a second prognosis about the FRT. The contractor contacts the OBI again half an hour before they expect the definitive end of the failure recovery process and when they have reached this end.

-*Process step 7*: The Trdl explains the situation to the BO. During this conversation the Trdl, possibly again, explains the location and gives a description of the failure form. The Trdl states whether they are dealing with an urgent failure or not.

-*Process step 8:* This step comprises the feedback the Trdl receives from the BO concerning the situation. The corresponding "Trein Incident Scenario" (TIS) or train incident scenario is chosen.

A TIS is a standard scenario corresponding with certain characteristics of the failure, a first prognosis and actions which have to take place.

2.2. Failure Recovery System - "Meldkamer Spoor"

Experiment MKS is a project currently (September 2015) ran in The Hague and will expand to the rest of the Netherlands in the coming years. It started at the beginning of 2015 and aims to reach the following goals:

- Quicker recovery of calamities and disturbances;
- Improved prognosis making;
- Increased process efficiency;
- Preventive signalling and contesting of calamities and disturbances.

MKS has, so far, taken two major steps towards these goals in the form of two experiments. The first step taken is the extended role of the BO. The BO (under the name of MKS) will do the intake of failures instead of the OBI. Because of this, the two communication processes shown at Figure 2.1 are combined. This thus reduces the amount of calls the Trdl has to do from two to one. The introduction of the MKS makes the entire process less error-prone and resulted in a reduction of the intake time of 2:30 minutes. An overview of the process with the MKS-department can be seen in Figure 2.2.

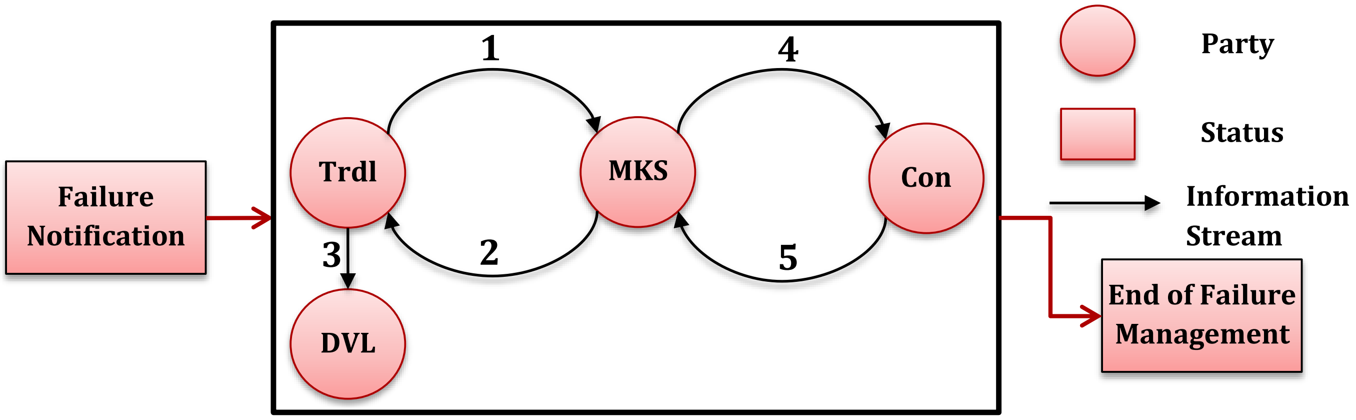


Figure 2.2: The Failure Recovery Process ProRail uses during the Experiment MKS.

The main differences between the processes of Figures 2.1 and 2.2 are the following:

- The BO and OBI are combined into the MKS. The departments BO and OBI are, for now, still available to support to MKS. The third parties will also still call the OBI;
- The second process step does now also include the determination of the timeslot for the contractor to repair the failure. This is determined by the MKS in consultation with the Trdl;
- The information which is already transferred in previous process steps (as explained in the points above) is, obviously, not transferred in later process steps.

The second big step experiment MKS has taken took place in July 2015. During this step an experiment is ran which includes the differentiation between two alarm lines: one for urgent calamities and disturbances and one for all the other notifications. This enables ProRail to handle these notifications differently in the future. The Trdl choses which of these two lines to call in a situation.

This research supports the experiment MKS in multiple facets. First, the FRT prognosis which follows from this research is delivered and used by the MKS (when it is expanded to the rest of The Netherlands) to base the length of the timeslots on. The FRM tool is also useful, in this case, for the second step of the experiment. The Trdl has to decide which alarm line to call in the future: urgent or non-urgent. This decision can be supported by the FRM tool.

2.3. Making the First Prognosis – Current Situation

From the current process steps explained in chapter 2.1, two steps may result in an indication of the expected FRT: step 2 (BO gives feedback to Trdl) and step 7 (OBI contacts contractor). Which of these steps actually results in the prognosis depends on the priority of the failure.

If the failure is a priority 2 (an UO failure), the prognosis is made by the BO in communication step 2. This is done based on the corresponding TIS. Every TIS comes with a prognosis about the FRT. These prognoses were formed some years ago by taking the mean of the duration between the arrival time of the mechanic and the time when the sign function recovery is given. An overview of the TIS's and their corresponding length is given in appendix C. As can be seen from the table in the appendix, there are five different TIS-types. These range from 1: disturbance to the train traffic to 5: discovery of a bomb. Usually, the higher the number, the more severe the calamity.

An important remark here is that the prognosis used for the TIS's includes the driving time of the contractor towards the failure place (the response time). However, the response time is not correlated with the failure type and can therefore be omitted from this prognosis evaluation without resulting in problems. Previous project research has determined the average response time at 43 minutes. Thus, this time can be subtracted from the TIS prognosis. Another remark concerning the TIS is that these are not specifically infra-failures, for instance: a fire next to the railroad. No infra-equipment should be repaired in this situation, thus the contractor does not play a role. These types of failures are not within the scope of this research. However, since this chapter focuses on the evaluation of the current prognosis method, this data can still be used for this purpose. During later analysis, these data are omitted.

When a failure is classified as a priority 5 failure (a DOT failure), an indication of the expected FRT is formed in communication step 7. This is displayed in the form of the length of the timeslot arranged with a contractor. However, this is done very roughly and not based on the FRT-expectation but on the train traffic circumstances. Usually, the begin- and the end-time of the timeslot are the times when the train traffic is absent or at its lowest point (which is usually during the night). Thus, the only prognosis which is actually based on the expected FRT and is therefore evaluated is the one for the TIS of UOs.

2.4. Performance

This chapter focuses on the current performance of the FRT prognosis making. These prognosis are judged based on their over- and under-estimation of the real FRT. Combining these two performance measures results in an overview how long the railway paths are closed while this isn't really necessary.

In order to measure the performance, data mining techniques are used as the base for the methodology to analyse the big SAP-database of ProRail. Generally, data mining consists out of five steps: problem statement, data collection, data pre-processing, model estimation and model interpretation (Figure 2.3). Using these steps is necessary to be able to efficiently and correctly deal with a large database like the one ProRail uses (Kantardzic, 2011).

In this case, a model is already present (see Chapter 2.3) but needs to be evaluated. Thus, the last two steps are not necessary. Only the first three steps of the data mining process are used. This results in a database with correct and consistent data. This database can be used to evaluate the performance of the first prognosis. This chapter thus first follows the first three steps, which are explained while they are applied.

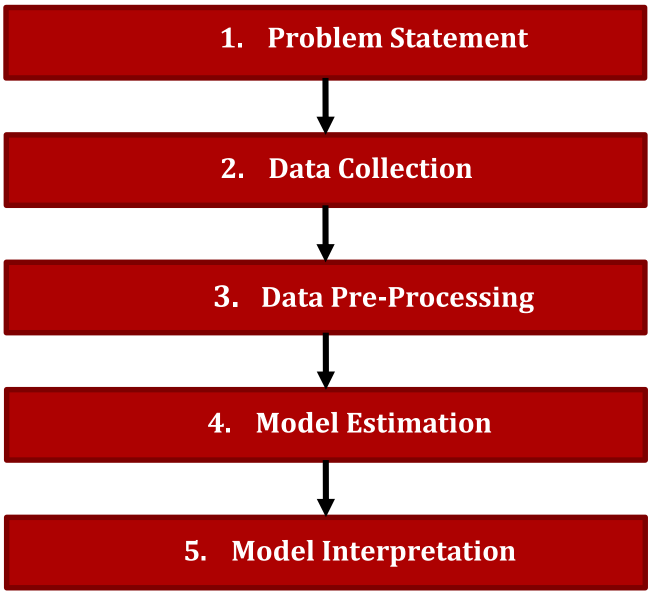


Figure 2.3: Steps in the Data Mining Process.

2.4.1. Problem Statement

This step identifies the problem which has to be solved. Both domain-specific knowledge and experience are necessary to form the problem statement (Kantardzic, 2011). This was done by conducting interviews with the database- and the asset management- experts to hear what was possible and necessary at ProRail. The problem (in the form of a question to be answered) which is solved in this chapter is the following:

How does the prognosis for the failure recovery time currently done by ProRail perform in practice?

2.4.2. Data Collection

According to Kantardzic (2014) there are two possible data-generation processes: designed experiment approach and observational approach. During a designed experiment approach the data-generation process is under control of the modeler (the one who executes the data mining process). As opposite to this, during the observational approach the modeler cannot influence the data-generation process.

This research deals with the observational approach as a data-generation type. The data is collected and being put into the SAP-system by the mechanics, the OBI and the VL as follows from the failure recovery steps in chapter 2.1. Therefore the modeler cannot influence this data-generation and has to deal with the data generated by other parties. It is important that this data-generation process is also used to generate the data used in the evaluation phase of the research in order to correctly evaluate the alternatives.

2.4.3. Data Pre-processing

Data mining techniques are valuable tools for decision support management. However, due to the huge databases, data in the real world tends to be noisy, incomplete and inconsistent (Zhang, Zhang, & Yang, 2003). Which is why data pre-processing is often necessary. ProRail is no exception here. Multiple pre-processing techniques exist but the most important ones are the following:

- *Data integration*. Data from multiple sources is merged into one coherent database.
- *Data cleaning.* Noise is removed, missing values are collected and inconsistencies in the data are corrected.
- *Data transformation.* Data are scaled to fall within a more suitable, smaller range suitable for data mining.
- *Data reduction.* The data is reduced by size by eliminating redundant features.

These pre-processing techniques are not mutually exclusive but can supplement each other in order to produce a reliable and workable database (Han, Kamber, & Pei, 2011).

After a quick analysis of the data-quality the conclusion was drawn to only use the data cleaning and data reduction techniques. Data integration is not necessary since the data necessary for this chapter of the research is all present in one database: the SAP-database. Data transformation techniques are not used because the data is already in the right scale or can relatively easy be transformed to the right format. First data reduction is used since data which is not used does not have to be cleaned.

Data reduction

During data reduction a huge database is reduced to a more manageable one. This new dataset is smaller in volume but still closely resembles the original database. Thus, mining in this database is more efficient while it gives the same results. Data reduction techniques can be classified into three types:

- *Numerosity reduction*. The original data volume is replaced by a smaller, but representable, data volume.
- *Data compression.* Transformations are used to obtain a compressed form of the original data.
- *Dimensionality reduction.* The number of random variables/attributes is reduced under consideration of not losing valuable information (Han, Kamber, & Pei, 2011).

Only the last data reduction technique is used. The first two techniques are relatively hard and time-consuming techniques that only pay off for extremely big databases, this is not necessary.

The dimensionality reduction is used by applying multiple filters on the data based on information acquired by data and failure recovery experts. These indicate why some of the data variables should be omitted or why certain attributes are not necessary. An overview of the applied filters can be seen in Table 2.1. This leads to a reduction from about 40.000 failures to about 2.500 failures.

Table 2.1: Filters applied to the Database for Data Reduction.

Filter	Motivation
Only data from October 2014 till September 2015	Only the data of the last year is used. Multiple failure recovery improvement programs are run at ProRail. Older data is therefore not representable.
Priority level one and two only	These are the urgent failures which are done as quickly as possible without a time-appointment. Therefore this data is not contaminated by non-value adding activities which tend to be present in other priority types ("smoking a cigarette, walking slow between failures etc."). This thus shows the possibilities for the FRT.
TAO: Yes	TAO ("Trein-aantastende onregelmatigheid" or train affecting incident) means that it influences the train traffic. If a failure does not influence the train traffic it can easily be repaired during the night, therefore the decision is easy and should not be included in the decision support tool failures.
Failure recovery status one only	These are the function recoveries corresponding with full recoveries. The other failure recovery types are not considered for this research question.
Only attributes relevant for the research	All the attributes which are not related to the research, like the type of contract or the equipment id, are removed from the file.

Data cleaning

Data cleaning (also called data scrubbing or data cleansing) deals with removing errors and inconsistencies in the database in order to improve the quality. Problems with the data quality are present in every individual database in the form of misspellings and missing information. These quality issues transform into problems as soon as multiple databases have to be integrated or a database has to be analysed (Rahm, 2000).

Han et al. classified the errors into three different types of mistakes: missing values, noisy data and inconsistent data. These writers also introduced multiple methods to treat each of these mistakes. During this data cleaning process one of his methods per mistake type is chosen and applied on the database.

Data with *missing values/attributes* are simply deleted from the database. The percentage of missing values is relatively low compared with the total amount of data. Therefore deleting these does not significantly influence the research results.

Noisy data is data influenced by noise, a random error in the variable. These data points are usually indicated by being outliers with either extremely high or extremely low values on certain attributes. Since the amount of data is not that high (after data reduction and some cleaning) human inspection is used to spot the outliers. This is done by sorting the data based on its FRT. The description of the extremely high and extremely low data points are individually inspected and, if the reason for its value is a mistake, removed from the data base. Examples of the reasons for the mistakes found are: no start date (therefore the first of January 1900 was taken automatically) or a failure which is, immediately after intake, switched to a different intake number.

The *inconsistent data* entries are removed from the database as well. Human inspection based on keywords is used here. All the descriptions of the failures are checked for multiple words indicating that the intake is wrong or double. If this is the case, it is deleted or related to another intake. At this point, the database is ready for examination.

2.4.4. Model Performance Estimation

The main goal of this step is to evaluate the prognosis made at the first stage of the failure recovery process. Thus, the prognosis based on the first information ProRail receives from the Trdl. To be able to reach this, the database created and cleaned above was analysed. First, each failure was classified into one of the TIS's which the prognoses are based upon (see appendix C). This classification into TIS's is done based upon the initial classification that the VL gives to the failure and the technical system of the infra-equipment. If this is not sufficient, the initial description of the failure entered into the SAP-system is consulted. An overview of the percentage of failures classified into a certain TIS is shown in Figure 2.4.

The meaning of the top eight TIS is explained and illustrated in Table 2.2.

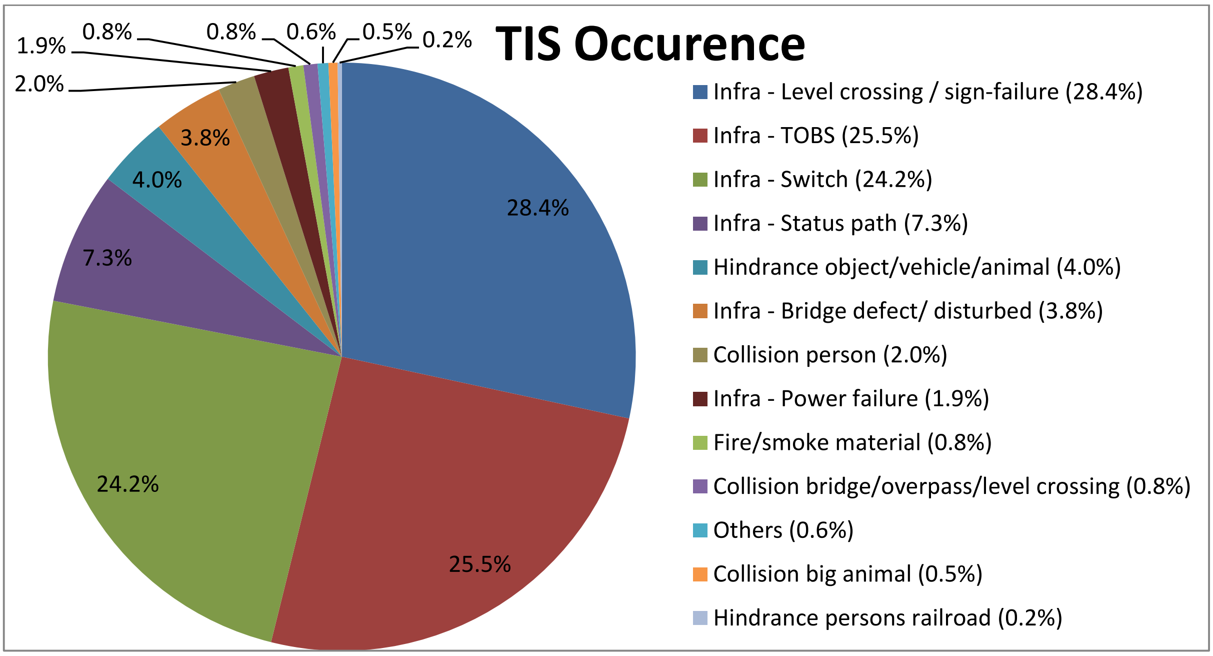
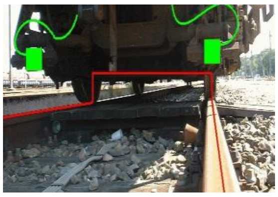
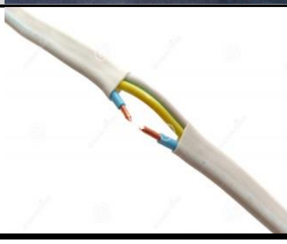


Figure 2.4: The Percentage of Occurrences for a TIS during the period October 2014 – September 2015.

An interesting conclusion from this chart is that three TIS's: level crossings/sign-failures, TOBS (the train-detection system which makes sure there is sufficient distance between trains fails) and switches, represent almost 80% of the total failures while ten TIS's represent almost 99% of all the failures.

The evaluation of the performance of the TIS prognosis is only done for the eight most occurring TIS's which represent 97.1% of the total amount of failures. This is done since the step between the ninth most occurring TIS (fire/smoke of material) with only eighteen failures and the eight most occurring TIS (infra – power failure) with 42 failures is significant. This higher amount of failure occurrences is necessary to draw reliable conclusions from the data. 42 Is just enough for this purpose while eighteen is doubtful.

Table 2.2: An Explanation and Illustration of the Top Eight most occurring TIS's.

Scenario	Explanation	Illustration
Infra-Level Crossing/Sign Failure	The level crossing ("overweg") or signs do not respond or are damaged (for instance by a lorry or an electronica failure).	
Infra-TOBS	The train detection system in the rails which makes sure there is sufficient distance between trains does not work properly. This causes, for instance, a level crossing to close while there is no train coming or paths not detecting anything when there is a train at that point: a "TOBS" or train vacancy detection failure.	
Infra-Switch	A switch ("wissel") does not respond or is damaged. For instance because it is frozen or there is a stone blocking it.	
Infra-Status Path	The status of the path is not or is suspected not to be sufficient to drive upon. For instance: a torn rails or eroded soil under the rails.	
Hindrance Object/Vehicle/ Animal	An object/vehicle/animal is on the track hindering the train traffic (thus, no collision between the two).	
Infra- Bridge defect/disturbed	A bridge used by trains to cross a river or road is defect or disturbed. This can be caused by for instance: an electronica failure or a lorry driving into it.	
Collision Person	A collision of the train with a person.	
Infra-Power Failure	Failure in the power supply. For instance in the catenary ("bovenleiding") or in a cable.	

These eight TIS's are evaluated next. Table 2.3 gives an overview of each TIS in order of occurrence frequency. It shows the given prognosis (thus excluding the response time of the contractor) and the actual average which follows from the data. The average is the average of the total function recovery time. This is the time between the registered arrival of the mechanic at the location and the sign: function recovered.

Table 2.3: Evaluation of the TIS Prognosis based on failure data from October 2014 till September 2015.

TIS	Prognosis (min)	Average (min)	Difference from Prognosis (min)
Level Crossing / Sign-failure	77	73	-4
TOBS	87	108	+21
Switch	77	80	+3
Status Path	167	125	-42
Hindrance object	67	55	-12
Bridge	57	75	+18
Collision with Person	102	74	-28
Power	77	121	+44

As can be seen from the table, the actual average differs from the given prognosis. However, the highest differences can be spotted at the TIS's with the least occurrences. This can therefore be blamed upon the relatively low amount of data for these scenarios. A few quite low or high values immediately results in a high difference of the average from the prognosis. For the three most occurring scenarios the difference is relatively small.

Table 2.4: The Amount of Data Points Smaller or Larger than the Prognosis and the Average Difference.

TIS	Prognosis (min)	Percent lower than Prognosis	Difference of Smaller Values	Percent higher than Prognosis	Difference of Larger Values
Level Crossing / Sign-failure	77	72.1%	47.3	27.9%	107.0
TOBS	87	62.5%	52.9	37.5%	143.8
Switch	77	67.4%	38.6	32.6%	91.4
Status Path	167	87.9%	107.1	12.1%	414.7
Hindrance object	67	75.1%	41.8	24.9%	77.9
Bridge	57	65.4%	31.4	34.6%	111.8
Collision with Person	102	81.1%	48.8	19.9%	54.2
Power	77	58.0%	52.0	42.0%	173.1

However, while the difference of the actual failure data and the data is not high, the amount of data points which are lower than the prognosis is very high, ranging from 58.0% to 87.9% compared to the data points higher than the prognosis. This is shown in Table 2.4. An important problem shown in Table 2.4 is that the percentage of data points higher (or lower) than the prognosis is not constant for different TIS's. This inconsistent performance of the prognosis makes it an unreliable source to count upon. The table also shows if a FRT is under- or over-estimated, by how much this is done on average. Huge average under- and over-estimations can be spotted here.

The high percentage lower than the prognosis can be illustrated by the data of the most occurring TIS in Figure 2.5. This picture shows the (annual) frequency in which each FRT for the TIS: infra- level crossing/sign-failures occurs based on FRT-intervals of fifteen minutes. The upper limit of each slot is shown at the x-axis. It also shows the cumulative of the percentage of the amount of failures in that FRT and the total amount of failures.

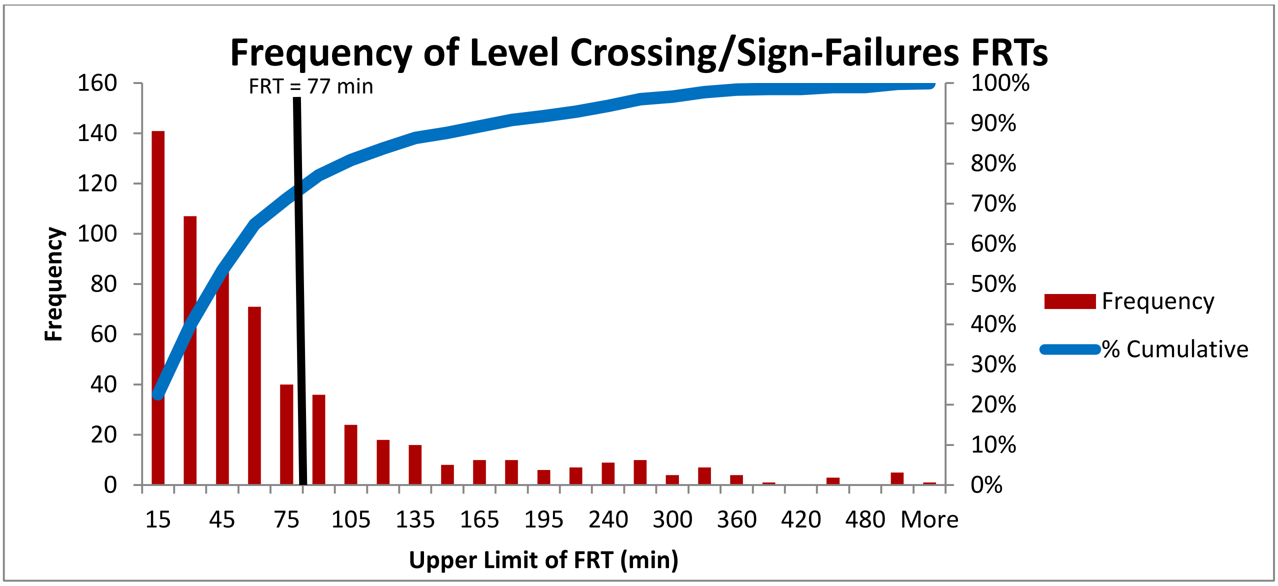


Figure 2.5: Frequency of the FRT for the Level Crossing / Sign Failures.

As can be seen from Figure 2.5, the data is highly skewed to the right. Thus, the "tail" of the probability density is to the right due to the relatively high amount of data with a really high value compared to the data with a really low value. This is a common and logical consequence of values with the unit time. For these units, the value can't be negative while there is no limit of the positive value. This is also a common phenomenon for failure recovery times since, if the failure is not found within a relatively short time period, the rarity of the failure cause increases and so also diagnosis time, repair complexity and the failure time (Wohl, 1982). The figure also shows that the frequency of data points lower than the prognosis tends to be significantly higher than the frequency of the data points higher than the prognosis. However, the prognosis for the TIS: infra – level crossing/sign-failures is still performing relatively well compared to other prognosis.

The scenario which has the largest difference between the amount of data points left and the data points right of the prognosis is the TIS: infra – status path. This scenario and its prognosis are shown in Figure 2.6. The amount of data points left of the prognosis is even higher in this case. There are also multiple data points which have a very large FRT. This indicates that the categorisation used for the TIS is not performing as it should perform.

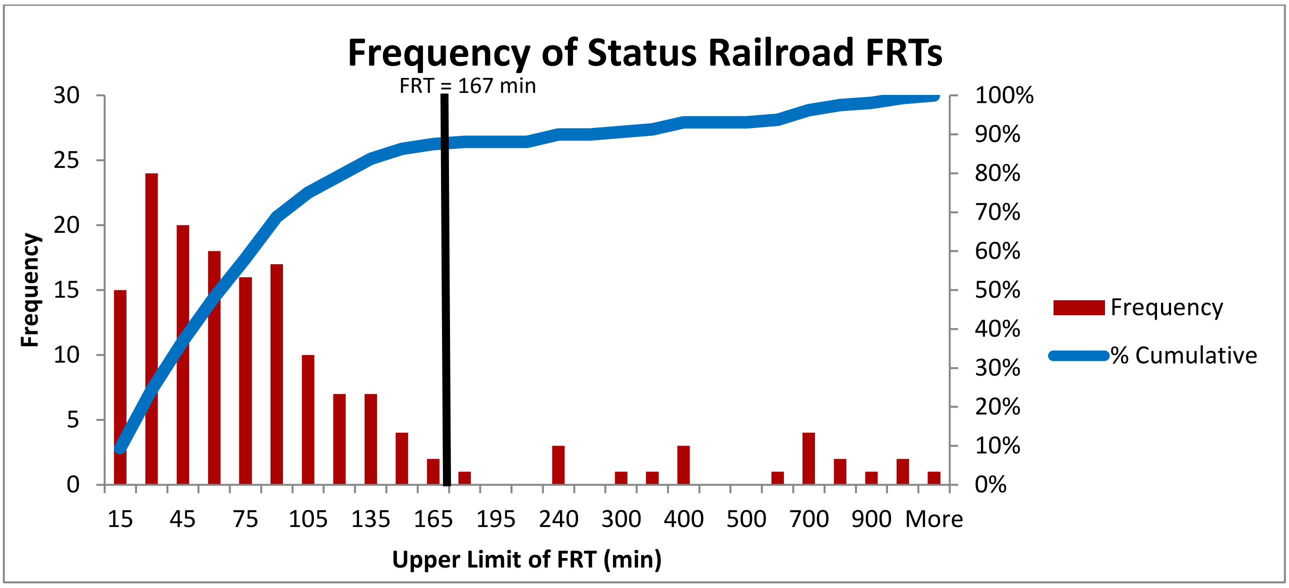


Figure 2.6: Frequency of the FRT for the Status Path Failures.

Both figures give reason to suspect a poor performance of the prognosis for TIS's based on the fact that data points are too spread out to work with one data point for the prognosis. This statement is reflected in Table 2.5 which shows the average, the standard deviation and the relative standard deviation: the coefficient of variation (the standard deviation divided by the average) of the most occurring TIS's.

Table 2.5: The Average and Standard Deviation of the Most Occurring TIS.

TIS	Average (min)	Standard Deviation	Coefficient of Variation
Level Crossing / Sign-failure	73	110.93	1.52
TOBS	108	156.53	1.44
Switch	80	100.31	1.25
Status Path	125	211.39	1.69
Hindrance object	55	69.84	1.27
Bridge	75	137.40	1.83
Collision with Person	74	51.31	0.69
Power	121	137.12	1.13

In most cases, the standard deviation is bigger than the average of the FRT. This is reflected in the coefficient of variation, which tends to be between one and two. This wide deviation from the average results in a poor performance of a prognosis based on the average and thus: one point only. It also indicates that a review of the failure classification in order to lower the standard deviation can be profitable.

2.5. Conclusion

From this chapter the conclusion is drawn that there are a large amount of steps and parties involved in the failure recovery process at ProRail, ranging from the Trdl to the contractor and depending on the type of failure. The research has to be conducted while involving and following advice of these parties while also taking the change to MKS into consideration. This change mainly combines the BO and the OBI into one department: MKS.

The chapter also shows that the prognosis which is currently used is not performing well. The current prognosis is made some years ago and merely shows the average time between arrival of the mechanic till failure recovery. This prognosis being outdated is shown by the performance of the prognosis for a TOBS. This prognosis is 24% too low in comparison with the actual average. For a failure of the failure form: status path this prognosis actually is 25% too high.

The poor performance is for a large part caused by the wide spread and thus, standard deviation of the data points. It is however also caused by the format of the prognosis which is a point estimate instead of an interval estimate. The estimate also is inconsistent in its under- and over-estimations: 72% of the level crossing failure FRTs are lower than the prognosis while this is 62% for a TOBS-failure. The format of the prognosis is therefore revised and also have a consistent performance for different failure forms. The chapter also identified the need for a review of the classification of the failures in order to reduce the large standard deviations.

3. Function Recovery Timeslot Length Prediction Methods

This chapter develops a set of useful methods for the determination of a suitable timeslot length for urgent failure recoveries. Therefore, the repair actions done by the mechanic are analysed first and thereby also the construction of the failure recovery time (FRT) out of individual times. This exposes the fixed and variable times within the FRT and results in a set of possibilities and restrictions. This is combined with the timeslot length determination methods to judge their suitability for ProRail.

3.1. FRT-Construction

The MTTR (Mean Time To Repair) of a system is the average time required to execute corrective maintenance. It is a useful parameter in the earlier plan- and design-stages of a system while still determining the location and thus accessibility of components in a system. It can be used to determine the life-cycle costs of systems (NASA, 2002) but also their availability. The availability of a system can be calculated by the well-known equation 1 (Sarkar & Chaudhuri, 1998).

$$Availability = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} \tag{1}$$

This equation gives two options to increase the availability of a system: either by increasing the MTTF (Mean Time To Failure) or by decreasing the MTTR. Traditionally, the focus is on increasing the MTTF. However, recent maintenance research has taken a different approach by focusing on the reduction of the MTTR (Fox, 2002) and has shown different circumstances where this focus results in a higher user-perceived availability (Song, Tobagus, Raymakers, & Fox, 2004). During this research the focus is also upon the MTTR, however, more upon the prediction of the MTTR times instead of its reduction. Since predicting the MTTR is necessary in order to reduce it, the same literature can be used as basis for both.

Maintenance literature identifies multiple steps usually included in the MTTR of corrective maintenance. The relevant steps for ProRail are the following:

1. Preparation time: this includes the steps necessary for an employee to prepare for a repair action.
2. Troubleshooting time: the investigation done to identify the exact failed part, its location and the failure cause. At ProRail, this is called the diagnosis time.
3. Spare part collection time: the amount of time necessary to gather a spare part. This includes the possible waiting time due to a shortage of a spare part.
4. Fault correction time: the actual repair. This is just called repair time at ProRail.
5. Evaluation time: the test if the repair is adequate or not.
6. Cleanup time: removal of dirt or other debris created by the repair action.
7. Return to service time: the time necessary to return the equipment to fully functioning (Cadwallader, 2012).

In the ideal situation the steps are simply followed from one till seven. However, loops back to earlier steps are also possible. For instance, when the evaluation of a failure recovery proves the recovery not to be sufficient in step five, a loop back to step two: a new diagnosis, is possible. Some of these steps are of particular importance in the railroad industry since they deal with severe safety regulations. This can be illustrated by an example of a repair of a particular failure form. In this example, the failure form treated is a section of a path wrongly detecting a train. Thus, a section of a railroad path which detects a train while there is no train there, thereby possibly not allowing other trains access to that section of the railroad path. This is caused by an electric circuit present between two sections of the path enabling an electrical current to run while this circuit normally isn't there (unless there is a train on both paths). The execution and time needed of the steps identified above for this example is shown in Table 3.1. The times are only a possibility, for another example they can be entirely different. Their order of magnitude is, however, relatively representable for other failure recoveries.

Table 3.1: An Example of the Failure Recovery of the Failure Form: Wrongly Detected Train in a Section.

Step	Actions	Time used to execute steps
1.	After some waiting for the Trdl to open the tracks, the mechanic gains access to the path and goes to the failed infra-equipment.	6 Minutes
2.	The mechanic identifies the failed infra-equipment as the ES-joint (electronic insulation-joint, this is used to detect a train by functioning as an isolation between individual rails. The train will add a way around this isolation enabling a current to run). The failure cause is identified as metal particles covering the ES-joint, removing its isolating function.	4 Minutes
3.	There is no spare part necessary and thus none collected.	0 Minutes
4.	The metal particles are removed and the ES-joint is covered with a protective coating.	6 Minutes
5.	The traffic control is called to test the switch and confirm it working again.	2 Minutes
6.	No cleaning is necessary in this case.	0 Minutes
7.	The mechanic exits the path and the traffic control opens the path again.	3 Minutes

Multiple failure recoveries have been analysed for these times. Based on this, multiple conclusions can be drawn. First of all, steps 3, 5 and 6 are usually either not even present or the time they take is negligible. The times they take are therefore included into step 4: the repair time, when calculations about the times are done. The focus is thus not upon these steps. Furthermore, the last step, the exit of the tracks, is relatively constant, small and barely influenced by the failure form, it is therefore omitted from the research. The first step represents the waiting time of the mechanic on the Trdl at ProRail. Just like the other times, the definition of ProRail is from now on used to indicate this time: the waiting time.

The three steps which are the most relevant to the total MTTR and are treated during the next chapters are therefore:

1. Waiting time
2. Diagnosis time
3. Repair time

These times significantly influence the total failure recovery time. Furthermore, the latter two, diagnosis and repair time, are also influenced by the different type of failure form (Cadwallader, 2012). These three are therefore interesting to further look into. The timeslot length determination methods applied on the data are for this reason focused upon these times. The times are illustrated in Figure 3.1. These are also accompanied with a variable indicating the exact point in time which can be used from now on. For instance: at time T_0 the mechanic arrives at the failure location while he actually starts diagnosing at time T_1.

Figure 3.1: The Individual Times during Failure Recovery

As explained in the first chapter, the result of the failure recovery times will be input for the decision support model for the VL (traffic control). The decision to be made is when to repair a certain failure. Relevant for this decision is the expected necessary length of a path withdrawal and thus: the time a mechanic actually needs the path. This is the time between T_1 and T_3. With the decision support model in mind, the time necessary for diagnosis and the time necessary for the actual repair can therefore be combined.

3.2. Timeslot Length Determination Methods

Chapter 3.1 identified the three most significant steps in the failure recovery process at ProRail: the waiting, diagnosis and repair time. It also argued that the latter two can be combined during this research. With a clear view of how the failure recovery at ProRail works and out of what elements the FRT consists, it is possible to determine a suitable list of FRT prediction methods. First, an initial list of possible methods is selected from literature. These methods are evaluated based on a list of criteria explained afterwards to end up with a definitive list which is applied on the data in the next chapter. The initial methods are selected based on their suitability to work with both numerical and nominal data (after cleaning) but also based on the time limit of the research. The latter does not allow complicated, time-consuming methods unless the less complicated methods prove not to be sufficient. Then more complicated methods might be tried if this seems profitable.

The methods are selected via literature research. The literature research does not limit itself to only repair time literature since general prediction methods also apply on the repair time predication and can therefore be useful. The methods found in the initial phase can be classified into four categories: regression analysis, time series smoothing methods, confidence intervals and other methods. Every method is explained and its most important advantages and disadvantages explained.

- Regression Analysis

Regression analysis tries to find the relationship between (several) independent variables and a dependent variable. In this case, the dependent variable is the total FRT time (or one of the three individual times) while the independent variable is unknown for now. Experiments to find suitable independent variable(s) are therefore necessary.

1. *(Multiple) linear regression*

Linear regression is one of the simplest regression analysis methods since it deals with only one single independent/explanatory variable. The model tries to fit a straight line through the set of data point in order to minimize the sum of the squared residuals (a residual is the difference between the data points and the line itself). The slope of the best-fitting line to the data points is called the correlation coefficient between the two variables (although slightly adjusted by the standard deviations of the points). An example of linear regression is shown in Figure 3.2 which shows the linear regression of the height and the body weight of people. As can be seen, the line fits relatively well, thus the squared residuals are low. In reality, a dependent variable rarely is only affected by only one independent variable. Instead, multiple factors affect the result. The extension to multiple independent variables is known as multiple linear regression.

Advantages: Easy method. Reliable predictions when a good independent variable is used.
Disadvantages: Only a linear relationship possible. There can be hidden independent variables (Brook, 2014).

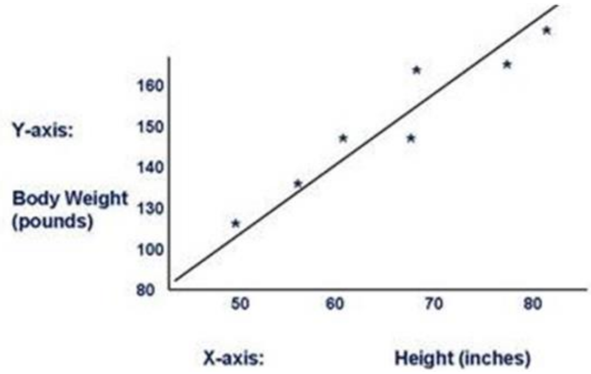


Figure 3.2: Linear Regression between Height and Body Weight.

2. *Non-linear regression*

This regression analysis method again consists of independent variables (explanatory variables) and one dependent variable (the response variable). However, now the model isn't explained by linearity but by another nonlinear function (for example: by an exponential, logarithmic or Guassian function). Some of these functions can be transferred to linear functions by linearization in order to simplify the model (Oosterbaan, 1994) but this is not always necessary (or easy).

Advantages: Suitable for every form of relationship. Reliable predictions when a good model is found.
Disadvantages: More difficult to develop and use the model. Not every independent variable can be tested due to time limitations (Brook, 2014).

- Time Series Smoothing

Smoothing methods are relatively simple methods to predict a future value based on previous data points. A *smoothing factor* is used to determine the role and the extent of the role that previous data points play. There are multiple methods of smoothing:

1. *Simple moving average*

During the simple moving average method, data points are smoothed out by calculating a moving average. This is, in this case, just the mean of the last k observations. The smaller the value of k is chosen, the smaller the smoothing effect will be and the more responsive the prediction is to recent changes in the data. The first k-1 predictions have to be based on less than k terms.

Advantages: Very simple to develop and use the model.
Disadvantages: Previous data points should be a reliable source of prediction for future data points in order for a good prediction performance (Goodell, 1963).

2. *Weighted moving average*

This is an updated version of the simple moving average. The same method is applied, however, each past observations now gets a weight. This weight determines the value of a past observation in predicting the new data. Usually, older data points get a low weight while the newer data points get high weights. Since recent data point usually have more value in predictions, this is an advantage over the simple moving average method.

Advantages: Simple to develop and use the model.
Disadvantages: Previous data points should be a reliable source of prediction for future data points in order for a good performance. Difficult to determine suitable weight values (Goodell, 1963).

3. *Exponential smoothing*

In the weighted moving average method, the past observations have constant weighting factors. During exponential smoothing, exponential window functions assign exponentially decreasing weights to these observations. Thus, the relevance of older data points decreases exponentially but this information is not entirely ignored in the prediction (like in the weighted moving average method). Exponential smoothing is illustrated by equation 2.

$$S_t = \alpha * X_t + (1 - \alpha) * S_{t-1} \qquad (2)$$

S_t Is the predicted value at time t. α is the smoothing factor and X_t the data point at time t. Thus, the data point at time t has a weight of α while the previous prediction has a weight of α-1. However, the previous prediction also constitutes out of its previous prediction and data point. When filled into formula 2 this automatically leads to a weight of α (1- α) for X_{t-1} and $(1- \alpha)^2$ for the prediction two time units ago. The older data points thus automatically gain less weight by this formula. (Goodell, 1963)

Advantages: Simple to develop and use the model. No historic data, and thus information, is deleted.
Disadvantages: Previous data points should be a reliable source of prediction for future data points in order for a good performance. Difficult to determine a suitable value for α (Goodell, 1963).

- Confidence Intervals

A confidence interval is an estimated interval about a certain to be predicted parameter. It is an old statistical tool first introduced by Neyman (1977) that has been successfully used over the last decennia to determine these intervals. The interval is determined based on a specific sample and thus differs per sample. The frequency of inclusion of the estimated parameter in the determined interval depends on the confidence level/coefficient. Thus, the higher this confidence level the higher the amount of intervals that contain the true value of the estimated parameter when a constant number of intervals are plotted. This higher amount is accomplished by increasing the width of the interval. Confidence intervals are a powerful tool when dealing with an uncertain parameter (Neyman, 1937). When applied to the necessary timeslot length problem, the unknown parameter is the average recovery time. However, only the maximum boundary of the interval is important for ProRail since the value which should be taken for the lower border is already known: zero minutes. Since the FRTs with a value close to zero are the most occurring it is necessary to, at least, add these most occurring values to the confidence interval. Especially since lower than zero is not possible. So, only the maximum boundary of the confidence interval should be determined. There are different methods for this:

1. *By the probability distribution*

The most generally used confidence interval determination method is via the probability distribution of data points. Analysing the data points via statistical tools (for example: histograms and goodness-of-fit tests) should be used to estimate and verify the probability distribution. Via the estimation of the parameters of the distribution a confidence interval can be created (Neyman, 1937). Since the exact application of this depends on the distribution of the data points, this is explained more extensively when applied.

Advantages: Can be applied on every set of data points. Accurateness of prediction is known. Probability distribution type indicates data point behaviour within the interval.
Disadvantages: Interval can become large in order to maintain reliability of prediction.

2. *By the values of the data points.*

Another method to determine this interval (although, theoretically, not a confidence interval) is by looking at the current, sorted data points and selecting the point which cumulative percentage of total data points is closest to the confidence level. Thus, if the necessary interval is 90 percent of the data and there are 200 data points. The data point with the 180th lowest FRT is the maximum interval boundary. This method is simpler than the confidence interval by probability distribution, however, less accurate when the amount of data is low (leading to large gaps between data points).

Advantages: Easy method to apply. Can be applied on every set of data points. Accurate prediction model with a large amount of data points.
Disadvantages: Interval can become large in order to maintain reliability of prediction. Not reliable when only a small amount of data points are present. Behaviour of data points within the interval is not known.

- Other methods

Besides the regression analysis, smoothing methods and confidence intervals there are multiple other possibilities to determine the necessary length of the timeslot.

1. *Neural networks*

A neural network is a computer program applying pattern detection and learning algorithms to build predictive programs. There is not one method which encompasses the neural network method. However, in general, predictor values are the input of one process. Calculations are done on these, resulting in the input for a next process, and so on. Eventually this results in an estimator. The use of neural networks is a powerful but complex tool suitable to be used for multiple purposes ranging from predictions and clustering to making of analysis.

Advantages: Accurate prediction method. Applicable to different problem types.
Disadvantages: Hard to use and develop (Berson, Smith, & Thearling, 2000).

2. *Nearest neighbour*

The nearest neighbour prediction method makes use of the idea that to determine the prediction value for a certain situation, you should compare this situation with historic situations. The actual value of a relatively similar historic situation ("the neighbour") should be used as the prediction value of the real-time situation.

Advantages: Simple to use. Improves automatically when the database gets larger.
Disadvantages: Requires a large database to begin with. Needs a good determination of rules which determine when two situations are neighbours (Berson, Smith, & Thearling, 2000).

3. *Prediction by expert*

An easy method capable of producing a prediction of the FRT is by letting the expert/mechanic, the actual person doing the repairs, predict its duration. For each failure form multiple mechanics can explain what they think about the necessary amount of time to repair it. Dealing with optimism/ pessimism of different mechanics can be done with, for instance, PERT (Program Evaluation and Review Technique). PERT estimates the expected time necessary (T) by taking a weighted average of the optimistic (O), pessimistic (P) and most likely estimate (M) via equation 3.

$$T = \frac{(O + 4 * M + P)}{6} \tag{3}$$

Advantages: Simple to develop and apply prediction method.
Disadvantages: Over-/under-estimation of themselves is a risk for the accuracy of the method. Mechanics may be reluctant to share information with ProRail since they are working for another company and ProRail is a customer of them.

4. *Minimum recovery time*

From the customers of ProRail, the railway undertakings, the request came for a minimum excepted FRT (function recovery time). This results in a minimum time the railway undertakings can expect the tracks to be in the non-perfect state in which they are at that point. However, since most of the function recoveries occur within the first few minutes (see Figures 2.5 and 2.6) a minimum time for the FRT will be zero minutes or, including response time (the time the mechanic has to drive towards the failure location, 43 minutes). Since this research does not look at the response time, this method is not suitable.

3.3. Performance of the Timeslot Length Determination Methods

For the selection of suitable methods for application and testing out of the long list of previously stated possibilities certain requirements of the method are determined. This was done based on the initial problem description delivered and interviews with involved persons. Based on these requirements a list of methods can be defined which can be applied on the data and evaluated afterwards to look for the best performing method. The suitability of each method is judged based on multiple criteria developed as explained above. These criteria closely resemble the criteria often used in the performance measurement of vehicle routing[4] heuristics: simplicity, speed, accuracy and flexibility (Cordeau, Gendreau, Laporte, Potvin, & Semet, 2002). However,

[4] Vehicle routing problems deal with the problem to determine the optimal route for a vehicle fleet to follow in order to deliver a product to a certain set of customers.

these are slightly adjusted to be able to use them in this specific situation. Besides this, the time limit of the research is also taken into consideration, thereby adding the criterion development simplicity.

The following requirements are important for the timeslot length method:

1. Simplicity: The method used for the necessary timeslot length determination should be suitable for practical usage. Thus, it should be understandable for an outsider (an outsider being a person with an average education).
2. Speed: The necessary length of the timeslot is needed by a Trdl (movement inspector) in order to judge whether to do the failure recovery now or later and to determine the length of the timeslot to reserve for this if done later. Since a Trdl wants to determine this rapidly it is necessary that the method should return results rapidly.
3. Accuracy: The method should be relatively accurate in its prediction. It should therefore be capable of displaying multiple probability distributions and also be capable of dealing with outliers.
4. Development flexibility: When new information about a failure form comes in, it should be possible to adjust the method or values used in the method based on this new information.
5. Development simplicity: The method should be able to be developed within a few weeks. This permits it to be tested throughout this research.

Every previously explained methods performance can be assessed based on these criteria. Since the methods are tested more accurately on their performance combined with the data later on in this report, only a rough performance estimation on each criterion is necessary at this point. The weights of the criteria are therefore equal. This performance estimation on each criterion is either positive, negative or neutral. When a method scores negative on a criterion it is evaluated whether the method is still suitable for this situation. The meaning of each of these performances is shown in Table 3.2 while the scores on all criteria per method and the total sum of them are shown in Table 3.3. These scores are determined based on literature (see explanation of the methods above) combined with the situation at ProRail. Not every score can be explained by Table 3.2, when this is not the case, an explanation is given underneath Table 3.3. This is done per criteria.

Table 3.2: The Meaning of a Score given for a Method on a Certain Criterion.

Criteria\Score	+	+/-	-
1. Simplicity	Easy to understand for an outsider.	Medium difficulty to understand for an outsider.	Hard to understand for an outsider.
2. Speed	Figure can be used to display prediction	Calculation should be done to display prediction.	Program should be used/ long calculation should be done.
3. Accurate	Model can explain all data points.	Model can explain most of the data points	Model does not explain most of the data points.
4. Development flexibility	Improvement can be implemented easily.	Medium difficulty to implement improvements.	Hard to implement improvements.
5. Development simplicity	Can be developed in a couple of hours.	Can be developed in a couple of days.	Needs more than a week to be developed.

Table 3.3: The Scores of the Methods per Criteria.

Method\Criteria	1	2	3	4	5	Sum
Linear Regression	+	+/-	+/-	+/-	+/-	+
Non-Linear Regression	+	+/-	+	+/-	+/-	++
Moving Average	+	+/-	-	+/-	+	+
Weighted Moving Average	+	+/-	-	+/-	+/-	0
Exponential Smoothing	+	+/-	-	+/-	+/-	0
Confidence Interval via Probability Distribution	+	+	+	+/-	+/-	+++
Confidence Interval via Data Points	+	+	+/-	+/-	+	+++
Neural Networks	-	-	+	-	-	---
Nearest Neighbour	+	+/-	+/-	+/-	+/-	+
Prediction by Expert	+	+	-	+/-	+	++

Simplicity: The neural networks are very hard to explain to outsiders, even experts have difficulties understanding how they work (Berson, Smith, & Thearling, 2000). The other methods are relatively well-known methods and therefore not hard to understand.

Speed: Only the neural networks score negative here. This is due to the fact that this method requires a format-change to the input but also the use of an extensive calculation by a software program, thus this takes a while. The other methods merely require simple calculations or looks at figures/intervals. This does not take a lot of time and thus quickly gives results.

Accuracy: The smoothing methods score negative on this criteria since they require a data point being influenced by its predecessor(s) in order to be accurate. However, since it is not the same infra-equipment which failed, these are spread out throughout the entire land and repaired by different contractors, this is not the case. The prediction by an expert is also not accurate since the amount of predictions made by experts is not high (due to the position of their company in the supply chain and the amount of experts/mechanics) and therefore not reliable. The other methods are all relatively accurate with positive exceptions for neural networks, confidence intervals by probability distribution and non-linear regression. However, for non-linear regression this depends upon the suitability of the independent variables. If good independent variables can be found, this method performs well and is therefore worth a try.

Development flexibility: Only neural networks are hard to improve. They require lots of new calculations and a change in the software which requires a lot of time and effort. The other methods merely require another formula or interval to be developed.

Development simplicity: Moving average and prediction by mechanics are relatively easy to develop. They require respectively a short look into the data or a mail/interview with a mechanic, this can relatively easily be done. Only the neural networks take a long time to be developed. They require a lot of research into the variables which play a role and the interrelationships of these. Besides this, for a practical application, software should be developed.

From Table 3.3 it can be seen that there are multiple methods with a positive score. The definitive selection can be determined by comparing the scores between methods within a category. Within the regression analysis category, the non-linear regression outperformed linear regression and scored positive. Non-linear regression is therefore applied. Only moving average scored positive from the smoothing methods. However, all of these methods are not accurate due to the fact that regression with the previous data point is necessary. Since this is not present, these methods are not suitable and applied. Both the confidence interval methods scored well. Since the confidence interval via data points only scores well when there are enough data points, the confidence interval via probability distribution-method is preferred. From the other methods, nearest neighbour and prediction by the expert all had positive scores and are therefore tested.

3.4. Conclusion

This chapter identifies individual times which play a role in the total FRT. We find that the FRT is constructed out of multiple individual times which can be combined into three times: the waiting, diagnosis and the repair time. We show that, for the decision support model, the most relevant length is the total time necessary for a path withdrawal and therefore the combination of the diagnosis and repair time. It should be kept in mind that the waiting time is excluded from this.

Furthermore, multiple possible FRT determination methods are identified. These are scored on multiple criteria and the best methods are selected for testing in later chapters. These are the following methods:

- Non-linear regression;
- Confidences interval via probability distribution;
- Nearest neighbour;
- Prediction by expert.

The methods identified above are applied on the data set. In the theory testing step (chapter 5), the best performing method is selected.

4. Performance of the Timeslot Prediction Methods

The methods introduced in chapter three are applied in this chapter. First, the new failure forms are determined. This is followed by the determination of the waiting time and an application of the methods identified in chapter three on the combination of the diagnosis and the repair time. This results in a theory formed per method. These theories can be tested in the next theory testing chapter to determine the best method and the relevant FRTs. The theories are only applied on the full failure recovery actions. The temporary repair durations are analyzed in Appendix I.

4.1. Failure Forms

Based on the analysis of the current situation done in chapter 2, we concluded that the current classification of the failure forms is not sufficient to base a prognosis on. Therefore, we attempt to form a new classification. This method might decrease the high standard deviation of the function recovery times based on failure forms as encountered in chapter 2. The classification was first done on a detailed level, thereby allowing the categories to be clustered later on if this is deemed more practical.

Classifying is based on the information available at the first failure notification. The information in this first notification is for instance: the location of the failure, urgency level or the time of the notification. Unfortunately, most of this information in the first failure notification is not related with the expected failure recovery time and is therefore ignored in the formation of failure form classification. The two fields which were relevant and therefore taken into account for this formation were the fields: "main-system" and "SAP-notification text".

The field "main-system" shows the infra-system in which the failed infra-object is initially classified. This is an initial, general classification system only based on eleven possible types of infra-objects instead of the way of failing. An uncontrollable switch is, for instance, classified in the main-system: switch. Since the way of failing is not shown in the main-system field, the field: "SAP-notification text" is also used.

"SAP-notification text long" is a text-field which shows the information about the failure the BO (back office) or the OBI (operational control centrum infra) entered manually into the SAP-system based on the information they receive from the Trdl (movement inspector). A typical example of this is: "Ddr: WL-1113 N.I.C." , with Ddr meaning the city Dordrecht, WL-1113 means "wissel" (switch) number 1113 and N.I.C. meaning "niet in controle"(not under control). Thus, a switch in Dordrecht can no longer be controlled by VL (traffic control). A lot of the entries have a standard format like this which enables failure form categories to be formed. The example above therefore falls into the category: switch NIC. However, there are also multiple non-standard formats used which had to be classified based on the previously stated field: main-system or by manually reading the text-fields. These were classified based on the most occurring way of failure.

It should be noted that now, unlike the TIS-scenario performance review, the failures without damage to the infra are omitted from the analysis. Since these do not need a repair, they do not fall into the scope of the research. The classification resulted in twenty-two categories, including the category: other failures. This last category includes the failures that are so specific that a new category suitable for them would not result in more than five failures in that category. Such a low amount of failures can't be used to base any useful statistically significant conclusion on.

An overview of the categories and the percentage of total failures they represent is shown in the pie-chart in Figure 4.1. The name of the failure form category is shown on the right. This name is accompanied with a number ranging from one to twenty-two. This number indicates the rank of occurrence and is later on used in figures/tables referring to the failure form accompanied with it to make these figures/tables a bit clearer. An explanation of every failure form is given in appendix D. This is also accompanied with the name every failure form would have at ProRail itself, thus in Dutch. As can be seen from the table, the three most occurring failures are: path TOBS (a train vacancy detection failure, the occupied path notification which is not correct in a section with just a path in it), switch NIC (a switch which can't be controlled) and level crossing disturbed (the level crossing barrier is closed for more than five minutes, automatically generating the sign: level crossing disturbed). These three represent more than 50% of the total amount of failures.

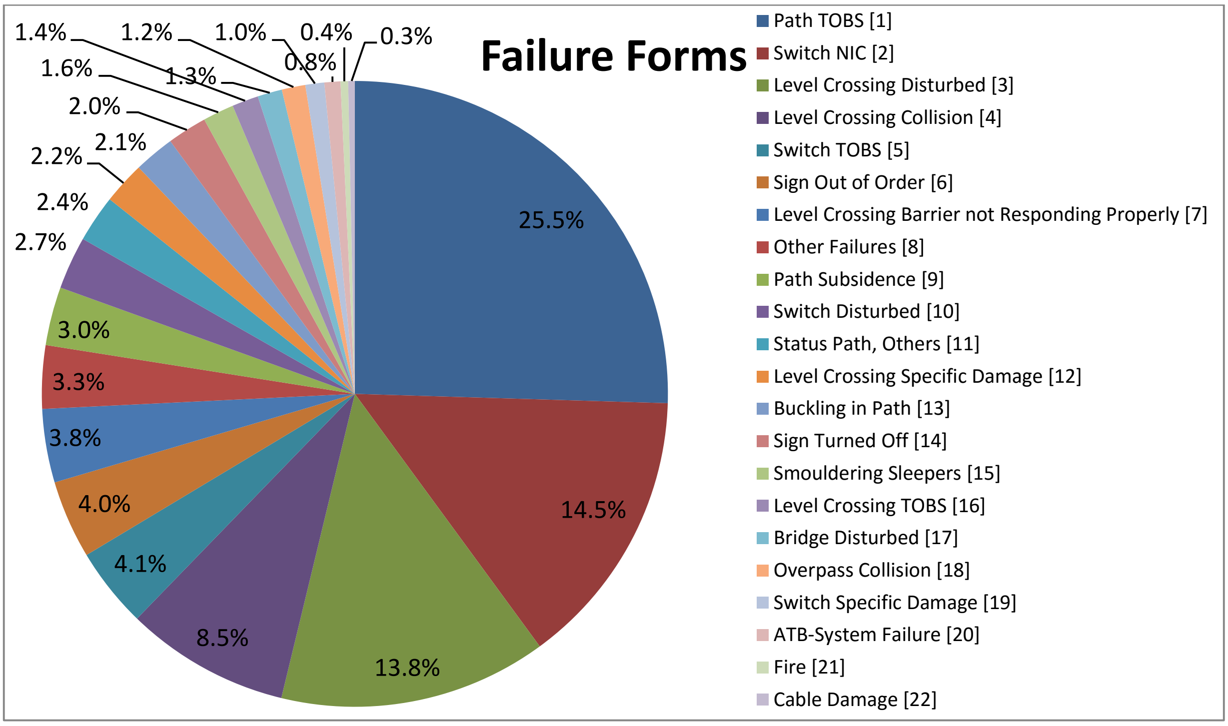


Figure 4.1: Overview of the Amount of Failures per Failure Form compared to the Total Amount of Failures.

This result was presented to the people who might eventually work with this: VL (the traffic control) and more specifically: the Trdls and DVLs of the railway-area: The Hague. The VL noted that their main interest is in the relatively regular failures: the failures with an occurrence higher than one percent. Therefore, failure forms 19 (switch, specific damage) till 22 (cable damage) are no longer considered in this research. Furthermore, VL noted that no clustering of the failure forms is required since this amount of failure forms is manageable.

4.2. Analysis Individual Times

As identified in chapter 3, the MTTR of urgent failures consists out of three individual times which are of significant impact on the total time: the waiting time of a mechanic, the time to diagnose the failure and the actual repair. While the combination of the diagnosis time and the repair time is the most relevant to the traffic control, the waiting time should not be accidentally included in this. However, there are some difficulties with the latter. Therefore, the waiting time analysis and a method to exclude it from the other times is done first. After this, the methods are applied on the diagnosis and repair time. This results in an overview how the data per time step was acquired and how the distribution of the individual time steps looks like. This can then later on be combined with the failure form.

4.2.1. Waiting time

The waiting time is the time between T_0, the arrival time of the mechanic at the train path containing the failure, and T_1, the time the mechanic can actually enter the path (see Figure 3.1). Duo to severe safety regulations in the railway industry, a mechanic is not allowed to enter the track without permission of the Trdl (the movement inspector). The Trdl only gives permission to the mechanic to enter the path as soon as there is a path withdrawal, the train traffic on that path is cancelled, which, in turn, depends on the severity and the impact of the failure on the train traffic and the time of the day. A failure only resulting in a train having to slow down at the failure point with no impact on surrounding paths and during rush hour results in more reluctance of a Trdl to call for a path withdrawal and thereby probably increases the waiting time of a mechanic. Not only the effects of a path withdrawal cause waiting time for the mechanic, this is also caused by the uncertainty of the expected arrival time. Since it takes some time for the mechanic to communicate with a Trdl to activate a path withdrawal, a Trdl has to know exactly when a mechanic arrives in order to call for a path withdrawal on time. Because, for example, traffic jams and traffic lights result in uncertainty in the arrival time and a too late call for a path withdrawal, some waiting time caused by this can be expected.

Thus, waiting time is caused by both the reluctance of a Trdl to call for a path withdrawal and the uncertainty in the arrival time of the mechanic. This suggests that the failure form barely plays a role in the expected waiting time only via the impact on the actual path (which usually is the same: unusable). This is worth investigating. However, first the data to do this have to be acquired. T_0 (see Figure 3.1) is already present in the SAP-database and was already used in combination with T_3 (the time a failure is recovered) in chapter 2 to determine the performance of the TIS prognosis. Unfortunately, a measurement of T_1 is not present in this database. Therefore, another database was used: the ISVL-database. ISVL is a database used by the traffic control (VL) and therefore contains the more logistically oriented data. In this case: the time path withdrawal is activated. The assumption was made that the mechanic enters the path as soon as a path withdrawal is activated, therefore the path withdrawal moment is suitable to be used for T_1.

Using the ISVL-database however, results in some difficulties. The most severe one is the low amount of data files with a filled in T_1 field. Most of the files did not fill this in since it is not mandatory to do this. The low amount of data obliges the research to determine a method to transfer the waiting time to the SAP-data without introducing unnecessary biases. Therefore, the data is analysed.

First the influence of the failure form is checked. This is done by classifying the waiting times. A low amount of categories is necessary here in order to cope with the low amount of data points. Since the impact on the train traffic can largely be derived from the location of the failure, the failure forms can be summarised into five different locations: sign (failure forms [6] and [14], switch (failure forms [2], [5] & [10]), level crossing (failure forms [3], [4], [7], [12] and [16]), path (failure forms [1], [9], [11], [13] and [15]) and others (failure forms [8], [17] & [18]). By doing this, the low amount of data is also not a problem while still being able to distinguish between different failures. Next, the waiting time data had to be cleaned. By manually checking the outliers, the most errors could be removed. Typical errors which were found in data recording happen when a mechanic could see the failure without the necessary path withdrawal. Thus, T_0 (the arrival time) is already registered while it takes a while before the path withdrawal because a mechanic already prepares the repair or goes away to pick up necessary equipment. Thus, the repair is already in progress while this still counts as waiting time.

With clean data, a t-test (or student-test) can be used to look for possible differences between the data sets accompanied with these failure locations. If there is a significant different, the individual failure forms can be tested later on to look at the details. A t-test is a statistical method to check the hypothesis that the means of two populations are equal while the standard deviation is unknown. A two-sided t-test is used. A two-sided test is useful to check for a difference in general, thus, without knowing if the mean of distribution x is larger than the mean of distribution y or that this mean is smaller. A significance level of 0.05 is used. The significance level is the criterion used for rejecting the null-hypothesis (the hypothesis that there is no difference between the distributions). If the test statistic (the p-value) is smaller than the significance level, there is enough evidence to reject the null-hypothesis and the outcome is statistically significant. 0.05 Is an accepted, often used, significance level representing a good cut-off point to reject the null- hypothesis (Craparo, 2007).

The p-value for every test of the locations compared to each other is shown in Table 4.1. The first row (t-test: sign vs. level crossing) shows the comparison between the waiting times of mechanics at the failure location: sign and the failure location: level crossing. Since the p-value is significantly above the 0.05 threshold, the null-hypothesis is not rejected. Therefore, there is no reason to assume that the waiting times of mechanics will differ between the failure locations: sign and level crossing. Every location is compared to every other location in the remainder of the table.

Table 4.1: The P-Value of the T-Tests between different Failure Locations.

	P-value
T-Test: Sign vs. Level Crossing	0.763
T-Test: Sign vs. Path	0.100
T-Test: Sign vs. Switch	0.680
T-Test: Sign vs. Remaining	0.422
T-Test: Level Crossing vs. Path	0.365
T-Test: Level Crossing vs. Switch	0.963
T-Test: Level Crossing vs. Remaining	0.541
T-Test: Path vs. Switch	0.053
T-Test: Path vs. Remaining	0,836
T-Test: Switch vs. Remaining	0.491

As can be seen from the table, all p-values are above the significance level of 0.05. Therefore, there is no reason to reject any of the null-hypotheses that there is a significant difference between the waiting times between failure locations. It is thus safe to assume the independence of the waiting time from the failure location and thus: failure form. Therefore, one waiting time determination for every failure form can be used.

The distribution of the waiting times into five minute intervals can be seen in Figure 4.3. The trendline indicates that the waiting times tend to follow an exponential distribution like the total failure recovery time in chapter 2. Thus, most of the waiting times are low but there are occasional large waiting times. 80 Percent of the waiting times are lower then 23 minutes while 90 percent of the mechanics can access the track after 35 minutes. Another interesting observation is the that 67 percent of the mechanics have acces to the track after just fifteen minutes.

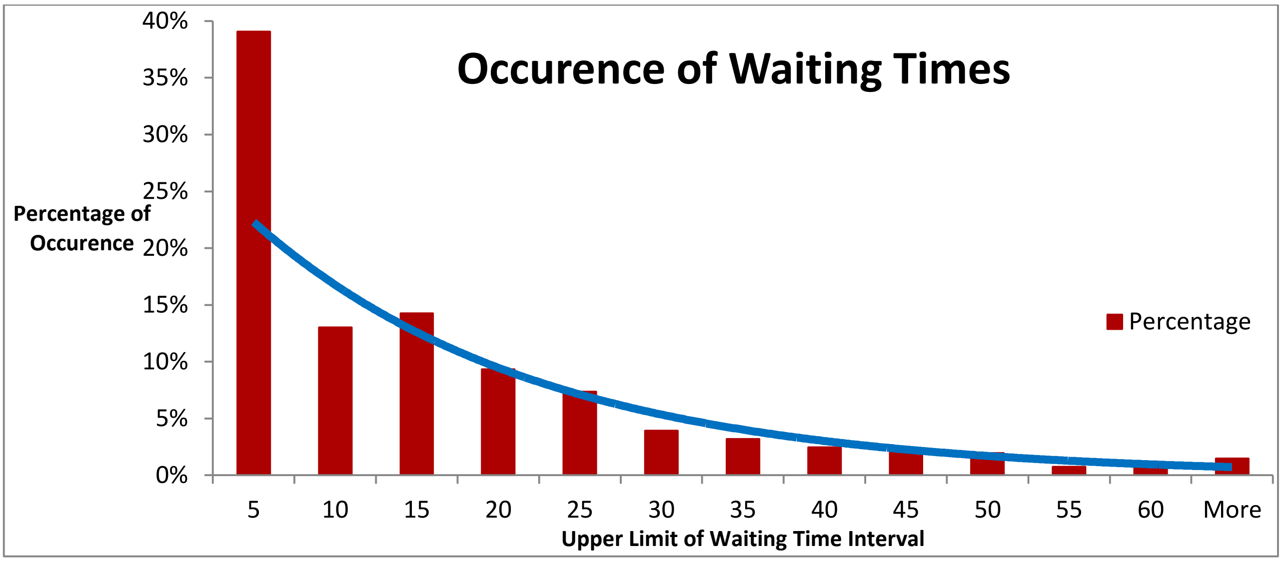


Figure 4.3: The Percentage of Occurrences of the Waiting Times of Mechanics.

The next problem with the waiting time which should be tackled before the final analysis is the transition of the data from the ISVL-database to the SAP-database. Since the data points from ISVL are not in accordance with the data points from SAP, this can't be done directly: subtracting the average waiting time of twenty minutes from a five minute failure recovery will result in a wrong representation of the data. A method to do this is developed with the help of Figure 4.4.

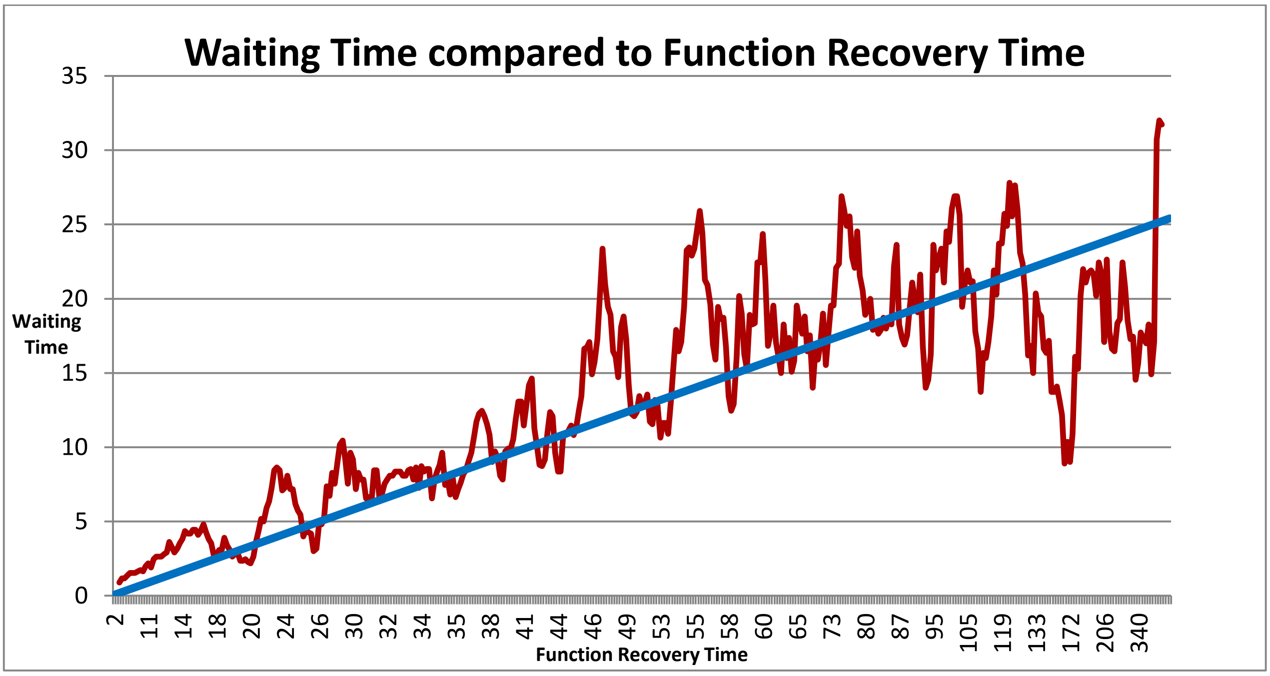


Figure 4.4: The Waiting Time compared to the Function Recovery Time.

Figure 4.4 shows the waiting time (T_0-T_1) compared to the total function recovery time (T_0-T_3) of individual data points (see Figure 3.1). In order to represent the trend in the data points, the waiting times were smoothed by taking the average of the waiting times of the five surrounding data points for every individual data point. This takes out the most extreme outliers while still representing some of the variance of the data points.

From Figure 4.4 it can be concluded that the waiting time follows a relatively constant line upwards in relation with the function recovery time. The usual standard deviation also seems to constantly increase while the function recovery time increases. This shows the constant link of the waiting time with the function recovery time. Of course, this is caused by the fact that a high waiting time is included in the function recovery time and, therefore, automatically increases the function recovery time. By taking totals, the averages can be calculated. On average, the waiting time represents 16.2 percent of the total function recovery time. Thus, when advancing to the diagnosis and the repair time, 16.2 percent of the failure recovery time should be subtracted from the time T_0 till T_3 for the waiting time.

4.2.2. Diagnosis and Repair Time

With the waiting time known, it is possible to determine the time necessary for diagnosis and repair of a failure (time T_1 till time T_3) for different failure forms. The classic experiment-theory is used as the methodology. Thus: first a theory is built based on a certain data set and this theory is tested afterwards on another data set. The data set used for the waiting time determination (the urgent failures from September 2014 till October 2015) is also used for the theory building part. The theory is tested upon a new data set in order to verify and possibly modify and enhance the theory. This also results in the best performing method (Colquitt & Zapata-Phelan, 2007). Four theories are formed based upon the four different methods identified in chapter three. The four methods applied are: prediction by expert (mechanic), confidence intervals, regression analysis and nearest neighbour. The theories built based on these methods are explained next.

Prediction by expert

The mechanic is the person actually repairing the failure and therefore the expert in this case. He works on behalf of the contractor and is located, with his colleague mechanics of the failure recovery team, in a specific area. For instance: the area Amsterdam, Weesp and Almere. When he receives a call from the Stoco (the "storingscoördinator" or failure coordinator) about a failure, he goes to the failure location, waits for the sign track withdrawal given by the Trdl ("treinddienstleider" or movement inspector), diagnoses the failure and repairs it. For this section multiple mechanics were interviewed and an entire day of the failure recovery team of Amsterdam, Weesp and Almere was witnessed on the 12th of November, 2015 in order to gain more insight into the failure recovery times (FRTs).

The prediction of the FRT done by a mechanic serves as a more detailed look into the length of the FRT and the influencing factors upon it. It is not possible to determine an accurate FRT prediction based on the failure form by the judgement of the mechanics since they state it is impossible for them to tell this without knowledge about the cause. However, multiple process-wise tips for this research per rough failure type can be noted. The mechanics were also capable of comparing the failure recovery lengths of general failure types against each other. This can later on be used as a verification method. These general failure types are the following: switch failure (failure forms [2] & [10] from Figure 4.1), section TOBS (failure forms [1], [5] and [16]), sign failures (failure forms [6] and [14]), level crossing failure (failure forms [3], [4], [7] and [12]) and an insufficient state of the path (failure forms [9], [11] and [13]). Remaining failures are not treated since their occurrence is that low and the definition that vague that a mechanic can’t give an accurate predication about these. Besides this, the process-wise tips only count for some of these failures and not for others due to the high amount of different failures. The problems and relative prediction of the FRT for each general failure type are:

- <u>Switch failure</u>. The mechanics noted that there are different switch types which directly influence the length of the diagnosis time. For instance: the detection whether a switch responds or not can with one type of switch easily be seen while another type requires the measurement of voltages of different power circuits. The type of switch thus influences the length of the diagnosis time. Another interesting remark the mechanic made is the reliance on switch part suppliers. When the cause of a switch failure is identified as a failure of a micro-switch (a part of the switch), the mechanic is not allowed to replace this (even though some of them are capable). Instead of this, they have to contact the supplier of this part who will replace it themselves. The response time of the supplier instantly forces them to do the remainder of the repair of the switch in the night.
- <u>Section TOBS</u>. The main influence on the repair time for a section TOBS is the location of the section. When the section is in front of the station the section is short and there usually are multiple curves and switches in the section (failure form [5]: switch TOBS). This increases the chance of metal particles being the cause of the failure due to increased wear of the track in a curve or switch. Metal particles ruin the function of the ES-joint (as explained in Table 3.1). However, since this cause is easy to repair (only cleaning) the repair time is low. When the section is not in front of the station, the cause is usually something else and relatively hard to find. Especially with a long section which is typically the case with failure form [1] (path TOBS) and not with failure form [16] (level crossing TOBS). The mechanics expected the FRT of failure forms 6 and 15 to be shorter than the FRT of a switch failure while the FRT of a section TOBS is longer.
- <u>Sign failure</u>. The FRT of the sign again varies a lot depending on the failure form. When the sign just turns off (failure form [14]), this usually indicated that a replacement of the lamp is necessary. This FRT is the lowest so far. However, other failure characteristics, for example: a flashing light or incapability of operating the sign (failure form [6], indicate a more difficult failure cause and result in a higher FRT comparable with the switch failure according to the mechanics.
- <u>Level crossing failure</u>. The mechanics remarked that the main influence on the FRT of a level crossing failure is whether or not to replace the level crossing barrier (failure form [4]). When this barrier has to be replaced they have to pick up a new barrier from Hilversum (the only depot). This significantly increased the FRT to the level of the section TOBS. If this is not necessary (failure forms [3], [7] and [12]) the FRT of a level crossing failure is at the same level as the switch failure.
- <u>Insufficient state of the track</u>. An insufficient state of the track (for instance: a buckling in the path or a path subsidence) always takes a lot of time to replace according to the mechanics. Not only because a lot of tools and machinery are necessary to replace the large failed parts but also because the “second level” has to be present for the repair of several failures falling into this category. The “second level” in this case is the leader workplace safety (LWB) who is responsible for the safety actions in the area with the train traffic withdrawal. He has to be present with, for instance, a path subsidence. This results in some extra travel time, safety measurements and time necessary to guarantee

the safety of the path after the repair is done. All of this results in the highest FRT for an insufficient state of the track according to the mechanics.

An overview of the FRTs according to the mechanics comparing different failure forms is shown in Figure 4.5.

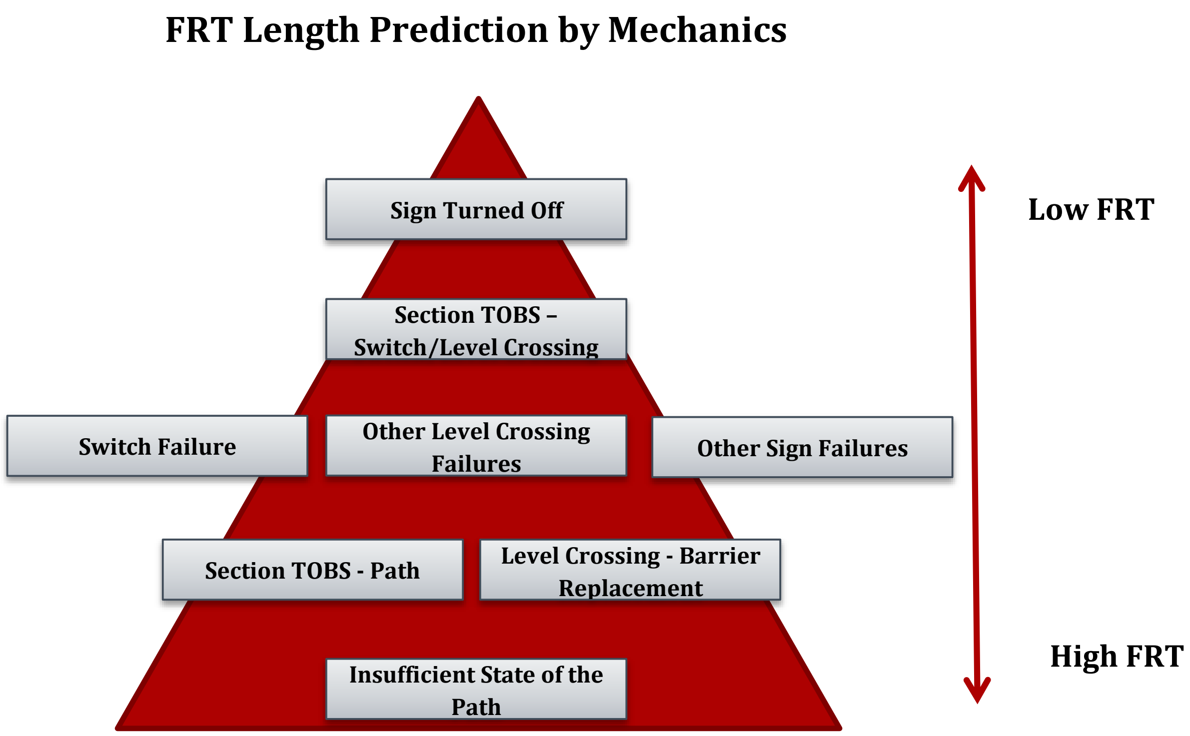


Figure 4.5: The Length of the FRT of Failure Forms compared to each other according to Mechanics.

Confidence intervals

A confidence interval is a good and reliable method to represent a data set. It deals with the uncertainty present in the data by giving an interval for the FRT instead of an average. In this case, the confidence interval is used to give an upper limit for the FRT in which, for instance, 80 percent of the FRTs are lying. This can be done by two methods: either by directly looking at the data points themselves and picking the data points corresponding with the 80 percent border or by developing the corresponding probability distributions and using these. As explained in chapter three, the last method is used. This method is not influenced by a lack of data points and therefore does not give an unfairly high boundary point.

Determining this confidence interval is done in a couple of steps. First, a suitable probability distribution and its corresponding parameters are picked for every failure form. This distribution can next be used to determine the interval boundaries representing a certain percentage of data points. Besides this, it can also be used the other way around by showing the amount of failures per failure form that can be solved within a certain amount of time.

These steps are done for every failure form as introduced in chapter 4.1. However, one more failure form is introduced or, to be precise, a previously used failure form is split up in this chapter. Also, some smaller, less occurring, failure forms are introduced and treated in Appendix F. The extra failure form treated in this chapter concerns the 'Diamond Switch with Double

Slips' or the 'Engels Wissel' in Dutch. It will be called diamond switch from now on. For a switch NIC the distinction is made between a normal switch and a diamond switch, which represent about 30% percent of all the switch failures: a significant amount. While a regular switch, shown in Figure 4.6, can only transfer a train from one track to another, a diamond switch, shown in Figure 4.7, offers the opportunity to directly cross a track. Since the last would usually cost two regular switches, diamond switches are usually present at busy train stations where space is limited. Whether you are dealing with a failed regular switched or with a failed diamond switch can immediately be seen when the first failure information about the failed infra-equipment is received by looking at the infra-equipment generating the failure notification.

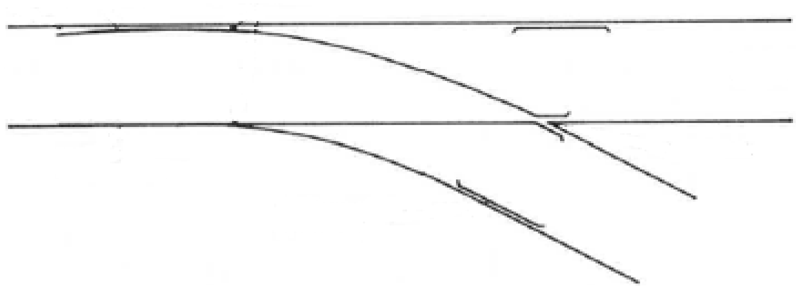


Figure 4.6: A Regular Switch

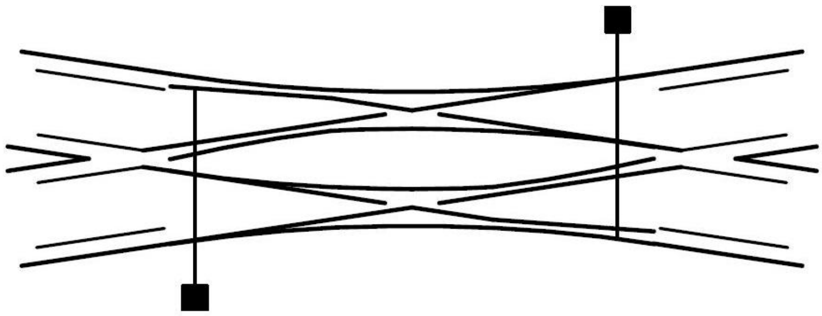


Figure 4.7: A Diamond Switch

The distinction between the regular and diamond switch is made at this point in the research because it does not influence anything but the repair time: the location and the size of the failed equipment, for instance, roughly stay the same. Besides this, some quick pre-testing showed that there seems to be a difference in repair time between the two. Other equipment distinctions don't seem to make a difference. Thus, also for the diamond switch a probability distribution is selected. The same goes for the smaller, less occurring, failure forms introduced in Appendix F. These are transferred to the appendix to keep the essay and results from overwhelming the reader. They also only represent a relatively small section of the failure forms. These latter failure forms are part of the overall failure forms: path TOBS and level crossing disturbed. By removing these smaller parts, the overall failure form becomes more reliable. Therefore, this step should not be ignored when looking for the most accurate (but slightly overwhelming) prediction.

Picking a suitable probability distribution can be a difficult process where, for instance, the mean, standard deviation, skewness and the graphical representation of a set of data points forms the basis for a possible probability distribution and its parameters. Whether this assumption is valid can be tested by the goodness of fit-test (Glover, Jenkins, & Doney, 2008).

Analysing the failure mode: level crossing disturbed by this method results in the presumption that the probability distribution of this failure form is exponentially distributed. This follows from the histogram (Figure 4.8) which roughly resembles an exponential distribution. Another indication are the almost similar values for the average and the standard deviation (respectively 47.6 and 44.9) and a level of skewness close to 2 (1.69). These are all indications of an exponential distribution (Ross, 2009).

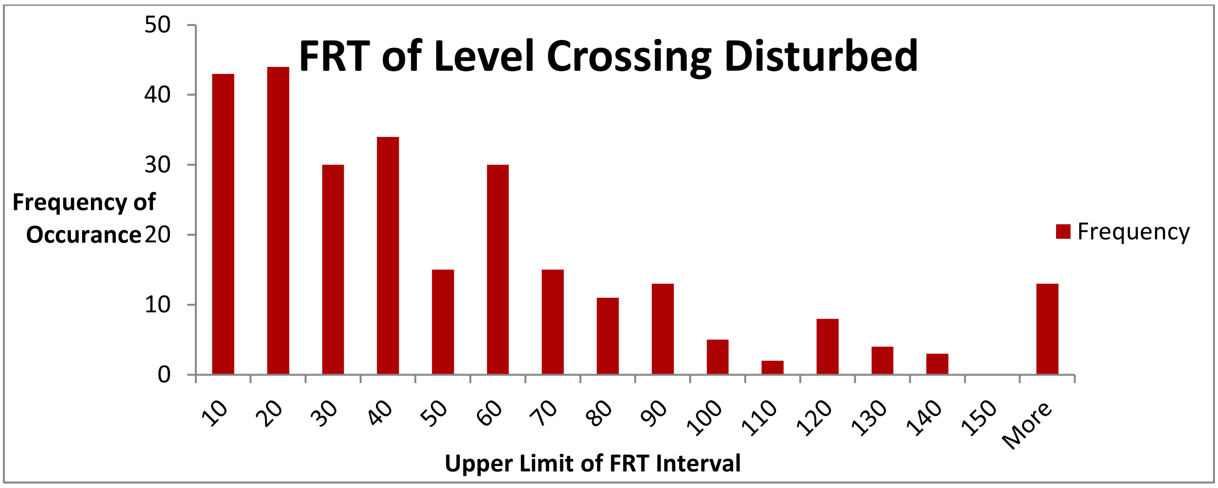


Figure 4.8: The Frequency of Occurrence for a FRT for the Failure Form: Level Crossing disturbed.

The next step would usually be to execute a Pearson's chi-square test in order to evaluate whether or not this assumption is correct. When the test-results are negative, another possible distribution has to be tried. It is easy to see that this method takes a lot of time. Therefore, the most likely distribution per failure form was automatically generated by the software program: Minitab. This program plots the data points and fits the best probability distribution to the points based on the minimal residuals (the distance between the actual data point and its assumed value according to the probability distribution). This results in the best possible distribution and corresponding parameters given in Table 4.3.

There are a couple of interesting observations to be made from this table. For instance:

- The shape parameter of the failure forms: switch NIC – regular, switch NIC – Diamond and switch disturbed are far above a value of one (see bold values). A Weibull and Gamma-distribution with a shape parameter of one indicates a constant repair rate over time. When the shape parameter is below one, this repair rate goes down over time. In this case, the repair rate increases. This means that the chance of repairing the failure in the next minute increases with time. This shows that the investigation time of the mechanic is useful, every minute he spends investigating the failure makes him more familiar with the failure and its possible solution.
- There are large differences between the scale parameters of the failure forms with a Weibull distribution. A higher scale parameters stretches the distribution of the recovery times to the right, this means: increases the expected failure recovery time. As can be seen from the table, sign failures tend to have a far lower scale parameter than, for instance, a collision with a level crossing.

Table 4.3: The Distributions and their Parameters for every Failure Form based on the Failures occurred from September 2014 till August 2015.

		Parameters			
Failure Form	**Assumed Distribution**	**Shape**	**Scale**	**Location**	**Threshold**
Path TOBS	Exponential	-	88.31	-	-
Switch NIC – Regular	Gamma	**2.22**	31.20	-	-
Switch NIC – Diamond	3-Parameter Weibull	**1.70**	65.58		-0.98
Level crossing disturbed	3-Parameter Weibull	1.03	48.90	-	-0.01
Level crossing collision	3-Parameter Weibull	1.14	77.30	-	-0.13
Switch TOBS	3-Parameter Weibull	1.05	67.09	-	0.99
Sign Out of Order	Exponential	-	56.36	-	-
Level crossing barrier not responding properly	3-Parameter Weibull	1.40	34.62	-	0.36
Other failures	3-Parameter Weibull	0.92	76.74	-	2.12
Path subsidence	Lognormal	-	0.99	4.36	-
Switch disturbed	3-Parameter Weibull	**1.53**	77.82	-	1.00
Status path, others	Exponential	-	71.73	-	-
Level crossing specific damage	3-Parameter Weibull	0.96	50.54	-	-1.03
Buckling in path	Lognormal	-	0.86	4.07	-
Sign turned off	Normal	-	30.88	20.53	-
Smouldering sleepers	Lognormal	-	0.68	2.78	-
Level crossing TOBS	3-Parameter Weibull	0.85	59.56	-	-1.44
Bridge disturbance	Exponential	-	47.44	-	-
Overpass collision	3-Parameter Weibull	1.18	16.22	-	1.97

From Table 4.3 it can be seen that five different probability distribution types describe the failure recovery time behavior of different failure forms. These are described next and their cumulative distribution functions (cdf)[5] are given:

- Exponential: An exponential distribution is a continuous distribution describing the time between continuously occurring events. An interesting property is that it is memoryless, thus, the current behaviour is not influenced by the past behaviour. The regular exponential distribution contains only a scale parameter (representing both the average and the standard deviation). This scale parameter is usually displayed in an exponential distribution with λ. The cdf of the exponential distribution is given in equation 4.

$$CDF(x) = 1 - e^{(-\lambda * x)} \qquad (4)$$

- Gamma: This 2-parameter distribution is also continuous and is frequently used to plot waiting times. Its shape factor k determines the shape of the distribution while its scale parameter θ determines the spread of the occurrences. The cdf of the gamma-distribution is shown in equation 5.

$$CDF(x) = \frac{\int_0^x t^{a-1} * e^{-t} dt}{\int_0^\infty t^{a-1} * e^{-t} dt} \qquad (5)$$

- 3-Parameter Weibull: The 3-parameter Weibull distribution can represent multiple other distributions. It is often used to model the time until a technical device fails. The

[5] The cumulative distribution function gives the probability for an occurrence smaller or equal to a certain value x based on the probability distribution.

parameters it uses are the scale (α), the parameter determining how spread out the distribution is, the shape (β), affecting the shape, and the transition parameter t_0 , the transition on the x-axis. The latter is called threshold in Minitab. The cdf of the 3-parameter Weibull distribution is given in equation 6.

$$CDF(x) = 1 - e^{-(\frac{x-t_0}{\alpha})^{\beta}} \qquad (6)$$

- Normal: An often occurring continuous probability distribution in which the probability of occurrence is symmetrically centered around the average. The normal distribution is accompanied with its average (location):μ and its standard deviation (scale): σ. The cdf of the normal distribution is shown in equation 7. Erf() represents the error function, the probability that a random variable with normal distribution (mean 0 and variance 0.5) falls into the range [−x,x].

$$CDF(x) = \frac{1}{2} * (1 + \mathrm{erf}(\frac{x-\mu}{\sigma*\sqrt{2}})) \qquad (7)$$

- Lognormal: The lognormal distribution is, again, a continuous probability distribution whose logarithm is normally distributed. The scale parameter (average):μ is given while the location parameter: σ represents its standard deviation (Walck, 2007). The cdf is given in equation 8.

$$CDF(x) = \frac{1}{2} + \frac{1}{2} * \mathrm{erf}(\frac{\ln(x) - \mu}{\sigma * \sqrt{2}}) \qquad (8)$$

Whether or not the assumed probability distributions from Table 4.3 are a correct fit for the data points can be checked with Pearson's chi-square test. This test compares the expected amount of data points in a certain interval to the actual amount of data points in that interval. Based on this, it computes the test statistic which can, in turn, be compared with the chi-square threshold value. When the test statistic is lower than the accompanied chi-square statistic, the distribution can safely be used. A more extensive explanation and a practical example of this method is given in Appendix E. In this appendix the failure form: level crossing collision is treated. This eventually leads to the conclusion that the assumed distribution (3-parameter Weibull) can safely be used for this failure form. An overview of all the test statistics and the accompanied chi-square value based on a significance level of 0.05 is given in Table 4.4. Besides this, the amount of clusters used in the test and the resulting degrees of freedom are given (see appendix E for an extensive explanation of these).

Table 4.4 shows that for every failure form the test statistic is lower than the accompanying chi-square value. Thus, it is safe to use the suggested distribution with a significance level of 0.05. However, in some cases the test statistic and the accompanying chi-square value are very close to each other. In these cases (for instance: path TOBS) the error margin is relatively low while the error margin of the assumed distribution of other cases is large. For instance, the failure form: sign out of order has a large difference between the test statistic and the chi-square value. Therefore, the assumed probability distribution can be accepted without a large amount of unexplainable variation.

Table 4.4: The Result of Pearson's Chi-Square Test for all the Failure Forms and their Suggested Distribution.

Failure Form	Assumed Distribution	Clusters	Degrees of Freedom	Test Statistic	Chi-Square Value
Path TOBS	Exponential	23	22	31.37	33.92
Switch NIC – Regular	Gamma	15	13	13.87	22.36
Switch NIC – Diamond	3-Parameter Weibull	9	6	7.30	12.59
Level crossing disturbed	3-Parameter Weibull	17	14	20.14	23.69
Level crossing collision	3-Parameter Weibull	13	10	7.62	18.31
Switch TOBS	3-Parameter Weibull	10	7	10.09	14.07
Sign Out of Order	Exponential	9	8	3.69	15.51
Level crossing barrier not responding properly	3-Parameter Weibull	9	6	8.44	12.59
Other failures	3-Parameter Weibull	8	5	7.98	11.07
Path subsidence	Lognormal	8	6	6.49	12.59
Switch disturbed	3-Parameter Weibull	8	5	7.53	11.07
Status path, others	Exponential	7	6	10.04	12.59
Level crossing specific damage	3-Parameter Weibull	7	4	2.09	9.49
Buckling in path	Lognormal	7	5	8.78	11.07
Sign turned off	Normal	7	5	1.74	11.07
Smouldering sleepers	Lognormal	6	4	3.26	9.49
Level crossing TOBS	3-Parameter Weibull	6	3	4.78	7.82
Bridge disturbance	Exponential	5	4	0.80	9.49
Overpass collision	3-Parameter Weibull	5	2	4.08	5.99

Based on the probability distributions the percentile intervals can be acquired. Which percentile intervals to acquire was determined with the help of some of the stakeholders of the process, in this case: a Trdl, DVL and the project owners of experiment MKS ("Meldkamer spoor"). The relevant inter vals to be determined are the 50%, 80%, 85% and the 90% intervals. Besides this, the percentage of failures per failure form capable of being repaired within 25, 50 and 75 minutes is relevant. 25 Minutes is a common length of a period given by a Trdl to a mechanic to do a first check of the condition of a failed infra-equipment. This period is based on the so-called: "four-train model" which is a common schedule of the amount of trains per path per hour: two "intercities" (trains stopping only at the major stations) and two "sprinters" (trains stopping at each station) per path per hour. Since the sprinters leave close (< five minutes) after the intercities from the station, there is a theoretical gap of 25 minutes between train traffic in front of the station. This is thus a policy based on minimal damage to train traffic. 50 And 75 minute periods are respectively two and three times a 25 minute period.

The resulting intervals and percentages are shown in Table 4.5.

Table 4.5: The Percentiles (in Minutes) and Percentages of Repaired Failures after certain Times per Failure Form based on its Probability Distribution.

	50% Repaired	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Path TOBS	45.2	122.6	148.0	183.8	39.6%	54.5%	65.7%
Switch NIC – Regular	43.2	82.9	95.2	112.0	35.2%	59.1%	76.1%
Switch NIC – Diamond	37.9	68.2	77.0	88.6	39.2%	66.0%	83.9%
Level crossing disturbed	18.3	58.0	71.4	90.2	59.6%	76.2%	86.1%
Level crossing collision	40.2	97.9	116.1	141.2	41.2%	58.7%	71.5%
Switch TOBS	30.3	85.1	103.0	128.0	48.6%	65.1%	76.5%
Sign Irregular	23.1	71.2	87.4	110.3	54.6%	70.9%	81.3%
Level crossing barrier not responding properly	16.3	28.8	34.9	43.0	76.2%	93.1%	98.3%
Other failures	33.4	106.9	132.0	167.9	46.8%	60.9%	71.0%
Path subsidence	62.7	161.7	229.0	260.7	28.3%	45.0%	57.3%
Switch disturbed	44.3	85.7	97.8	113.7	35.6%	57.7%	74.5%
Status path, others	33.7	95.9	116.6	145.7	46.2%	62.1%	73.2%
Level crossing specific damage	19.5	64.5	80.0	101.9	57.9%	73.8%	83.6%
Buckling in path	42.3	101.1	140.8	156.8	37.7%	58.1%	71.2%
Sign turned off	4.5	27.0	33.0	40.6	78.1%	94.4%	99.2%
Smouldering sleepers	6.1	8.9	15.4	18.7	93.4%	98.5%	99.6%
Level crossing TOBS	24.1	86.2	108.5	140.8	53.2%	67.4%	76.8%
Bridge disturbed	16.9	56.9	70.5	89.7	60.9%	76.9%	86.4%
Overpass collision	1.9	2.8	6.5	11.5	96.8%	99.7%	99.8%
Everthing (based on the data points)	31.9	96.3	120.4	163.4	46.0%	63.4%	76.2%

The table shows that, for instance, almost 60 percent of the level crossing disturbed failures are solved within 25 minutes. Besides this, 90.2 minutes is sufficient to repair 90 percent of these failures. It also shows the differences between the two types of switch NIC: a diamond switch is easier to repair. This last sounds counterintuitive since a diamond switch contains more components (see Figures 4.6 and 4.7), however, a closer look at the data shows the reason for this. Due to the high amount of components of the diamond switch, the failure reason is often located in one of these while the failure reason of a regular switch is relatively often another, harder to diagnose, reason.

The table can also be used to look at the successfulness of the 25-minute interval, meaning: the probability to solve a failure in the next 25-minute interval. These are shown in Table 4.6. This table shows that the probability to solve a failure in an interval when the last interval did not solve the failure usually goes down. This is logical since it indicates that the failure is hard to solve. However, the solve probability goes up for the switch failures. In these cases, the probability to solve a failure in the next interval is higher when this is the third 25-minute interval than when this is the first. It is therefore recommended not to postpone the failure recovery process when the first interval is not successful for a switch failure.

Table 4.6: The Probability to solve a Failure in the next 25-Minute Interval when this did not happen in the Intervals before.

	Interval 1	Interval 2	Interval 3
Path TOBS	39.6%	24.7%	24.6%
Switch NIC – Regular	35.2%	36.9%	41.6%
Switch NIC – Diamond	39.2%	44.1%	52.6%
Level crossing disturbed	59.6%	41.1%	41.6%
Level crossing collision	41.2%	29.8%	31.0%
Switch TOBS	48.6%	32.1%	32.7%
Sign Out of Order	54.6%	35.9%	35.7%
Level crossing barrier not responding properly	76.2%	71.0%	75.4%
Other failures	46.8%	26.5%	25.8%
Path subsidence	28.3%	23.3%	22.4%
Switch disturbed	35.6%	34.3%	39.7%
Status path, others	46.2%	29.6%	29.3%
Level crossing specific damage	57.9%	37.8%	37.4%
Buckling in path	37.7%	32.7%	31.3%
Sign turned off	78.1%	74.4%	85.7%
Smouldering sleepers	93.4%	77.3%	73.3%
Level crossing TOBS	53.2%	30.3%	28.8%
Bridge disturbed	60.9%	40.9%	41.1%
Overpass collision	96.8%	90.6%	33.3%

Regression Analysis

Regression analysis aims to find the relationship between one or more independent variables (the explanatory variables) and a dependent variable (the response variable). The dependent variable is, in this case, the function recovery time (FRT). Which independent variables are important to the dependent variable is subject of the regression analysis. Theoretically, every variable (for example: the failure location, the weather in The Netherlands or the performance of Manchester United last weekend) can influence the FRT and its influence should therefore be researched. However, due to obvious practical reasons, only variables which logically influence the duration of failure recovery are researched. One of these variables is already explained in the beginning of the essay: the failure form. This variable is that important that it is already used as a failure clustering method. The regression analysis will therefore be applied on individual failure forms. Another reason why the failure form is not a direct input factor but a clustering factor is because of the inter-factorial effects of the failure form. For instance: in the North-East of The Netherlands there are more easy-to-repair level crossings and signs. A short repair time in this region has therefore nothing to do with the region-effect but with the most occurring failure forms in the region. The failure forms are therefore isolated. The other probable relevant variables being researched in this chapter were selected by discussing this subject with experts. These are the following variables:

- The region the failure occurs in. ProRail divides The Netherlands into four different regions: South, North-east, Randstad (the region around the four largest cities in The Netherlands: Amsterdam, Rotterdam, The Hague and Utrecht) South and Randstad North. Every region has its specific characteristics, for example: amount of wind, amount of train path per square kilometer or amount of traffic jams. These characteristics of the region influence the failure causes and thereby the FRT. The more salty air in the north and west of The Netherlands, for example, increases the amount of rust on the infrastructure (Lewis & Schwartz, 2004) while a larger amount of traffic jams in the regions Randstad South and Randstad North in comparison with the other regions increases the time when a spare part has to be delivered from the warehouse.

- The type of contract a contractor has. At ProRail, a contractor can have two types of contracts: a PGO ("Prestatie Gericht Onderhoud" or performance focused maintenance)-contract of an OPC ("Onderhouds-Prestatie Contract" or maintenance-performance contract)-contract. With a PGO- contract the contractor gets paid based on the performance of the infrastructure while the contractor gets paid based on the performed activities with an OPC-contract. The former is a type of contract ProRail recently (2013) started working with and which motivates contractors to quickly resolve failures but also prevent failures from happening.
- The timing of the failure. This is a binary variable indicating whether or not a failure occurs, and is repaired, during rush-hour. Rush-hour is defined as the time between 7:00 and 9:00 and the time between 16:00 and 18:00. When a failure has to be repaired during this time the FRT is expected to be lower due to a higher relevancy of a rapid recovery of the paths. This possibility motivates the mechanics to work faster or the contractor to bring more personnel.
- Recent use of the grinding train. The passing of trains on train paths causes cracks to form in the steel paths. In order to prevent these cracks from becoming too large, a grinding train frequently (depending on the use intensity) treats a path. This train removes a small layer of steel from the path and thereby also removes the starting cracks. However, by removing this layer, metal particles are created. These are, theoretically, cleaned up by a magnet fastened on the train but occasionally still end up on the path. As explained before, these particles can cause failures in the form of failed joints. Since these type of failures are easy to repair (removal of the metal particles) failures caused by the grinding train have a low FRT. Therefore, this phenomenon in the form of a binary variable is included in the regression research. The variable is registered as one when the grinding train has been used on that specific piece of path in the last three days (according to experts, the failure would have occurred within three days) and is registered as zero otherwise
- Recent maintenance activities. A common reliability term is the "bathtub model". This model describes the higher failure rate of a technical system when it is new and at the end of its life. A new system tends to have some simple, infant failures while the later increase of the failure rate is due to wear of the system (Klutke, Kiessler, & Wortman, 2015). The implications on this situation of this model follow from the regular overnight maintenance activities. These activities aim to prevent failures by replacing or maintaining the infrastructure. However, like in the bathtub model, this can lead to failures during the next day. These failures are often easy to identify and repair, for example: a replaced fuse fell out. If there was any maintenance activity on the piece of infrastructure which failed, the variable is registered as one, otherwise as zero.

To summarize, one failure notification has multiple input factors which are tested. An example is given in Table 4.7. This table shows the values of the input factors for one of the Path TOBS failures of the year.

Table 4.7: The Input Factors for one of the Failures this Year of the Failure Form: Path TOBS.

Failure type	Path TOBS
Region	Randstad South
PGO-contract	1
Rush-hour	1
Grinding Train	0
Recent Maintenance	0

As can be seen in this example, there are no continuous variables in the regression analysis, only binary and attribute variables. This has an implication for the regression analysis: only linear regression will be possible. In order to prevent interactions between the different variables, the variable which definitely has an interaction effect is already removed from the analysis (the failure form). For the other variables, no logical explanation for an interaction effect can be found. It is therefore safe to assume that this interaction is not present. Because each variable can only be zero or one, it can only have two effects on the FRT instead of a constantly present but fluctuating influence depending on the variables value.

Acquiring the values of each variable at the time of the failure, again, requires some extensive database combinations. For example, the data about the grinding train is only accessible by a few persons responsible for this subject, while multiple format changes were also necessary to combine the data, once acquired, with the SAP-data.

To test which variables are good predictors of the FRT value, F-tests are executed. An F-test is, during regression analysis, a test about the fit of different linear models. It is capable of doing this for multiple coefficients at a time. Just like the t-test done earlier, this results in a p-value which can be compared to an earlier determined significance level. This will determine whether or not a specific variable influences the FRT (Markowski & Markowski, 1990).

The results of these tests are shown in table 4.8. This table displays the p-values per variable per failure form. Since not every variable took different values for every failure form, not every variable is included per failure form. For instance, there were no level crossing disturbed failures where there has recently passed a grinding train. In this case, the assumption is made that these variables are not related.

Furthermore, some of the failure forms were combined due to the lack of variation of variable values when the amount of data points becomes too low. Therefore, only the ten most occurring failures are done while the data of other failures will be added to data of the more occurring failure form most suitable according to the failure location (see section 4.2.1). For instance, the data of the failure form: "sign turned off" is added to the data of the failure form: "sign out of order", resulting in the same values for both failure forms. It is also important to note that the switch NIC now again refers to both the regular and the diamond switch type since, as explained earlier, influences of other factors apply the same on both switch types.

Interesting findings which can be observed from the table are the following

- The regions in which the failure occurs barely influences switch failures while it does influences the FRT for some other failures (see bold numbers). This can be explained by the placements of switches. These are always placed where paths come together, thus near stations and thus, switches are often in cities. Distances to spare parts in the mechanics car or testing opportunities, for instance, are therefore constant over different regions.

- The FRT of a switch-failure is influenced by the contract form (see italic number), however, other failure forms are not (or not at the same rate). This may be the case since switch failures are the most prestigious ones (due to their infamous public reputation) to perform well. Another explanation can be found in the bonus-structure of the contract. Performing well on a switch failure may be more attractive for a contractor than performing well on a level crossing failure because of the payments improvements accompanied with it.

Table 4.8: The P-Values of the F-Test executed on the Regression Variables per Failure Form.

			P-Values		
Failure Form	**Region**	**Rush-hour**	**PGO-contract**	**Grinding Train**	**Recent Maintenance**
Path TOBS	**0,02**	0.65	0.84	0.23	0.81
Switch NIC	0.32	0.34	*0.03*	0.65	0.36
Level crossing disturbed	0.08	0.08	0.14	-	0.94
Level crossing collision	0.08	0.13	0.34	-	0.66
Switch TOBS	0.59	0.26	0.19	0.95	-
Sign out of order	0.19	0.57	0.56	-	0.39
Level crossing barrier not responding properly	0.99	0.43	0.86	-	0.00
Other failures	0.45	0.24	0.53	-	0.67
Path subsidence	**0.04**	0.46	0.82	-	0.78
Status path. others	0.51	0.78	0.05	-	0.09
Switch disturbed	0.32	0.34	0.03	0.65	0.36
Level crossing specific damage	0.08	0.08	0.14	-	0.94
Buckling in path	0.51	0.78	0.05	-	0.09
Sign turned off	0.19	0.56	0.56	-	0.39
Smouldering sleepers	0.51	0.78	0.05	-	0.09
Level crossing TOBS	0.07	0.07	0.14	-	0.94
Bridge disturbed	0.51	0.78	0.05	-	0.09
Overpass collision	0.51	0.78	0.05	-	0.09

As can be seen in Table 4.8, the region plays a role in the FRT prediction per failure form. This is why the region is further partitioned into district control offices (Dutch: “VL-post”). A district control office is a building located near a major train station which takes care of the train tracks in that area. This area is smaller than one of the four regions used earlier. There are thirteen district control offices in The Netherlands. Each one of them belongs to one region. For instance, region South contains district control offices: Maastricht, Roosendaal and Eindhoven. This partitioning is done only for the failure forms for which the p-value for region in Table 4.8 is lower than 0.25. This value is low enough to indicate that the region and, probably, the district control office plays a role in the FRT.

Prediction the length of function recovery is also possible with regression analysis. In order to predict this, the software program Minitab was used again to speed up the process. This program applies an often used method to find the best fit of a linear function to the data points with the independent variables as input. By plotting different lines and determining the difference between the actual value and the expected value as predicted by the line, Minitab is capable of finding the line with the best fit to the data points. This is often revered to as: multiple linear regression (Mendehall, Sincich, & Boudreau, 1995).

The independent variables which are used as input factors are the ones in table 4.7 with a p-value below 0.25 (region thus portioned into the district control offices). This is a higher boundary than usually used, however, due to the low amount of data for some variables, sufficient to prove the influence of a factor. Besides this, if a p-value is high but still within the boundary of 0.25 it will barely have influence in the regression equation generated next.

The equations to predict the FRT per failure form are shown in Table 4.9. Sometimes, a value is not present where this one is expected to be present. This can be explained by the absence of certain failures forms in certain district control office areas. For instance, there are barely any level crossings in the area of district control office: Amsterdam. Thus, this variable is not present for corresponding failure forms. The equations tend to start with a constant, followed by the influences of the relevant input variables. These input variables always either take a value of one when present, or zero when absent. Thus, the example of Table 4.7, path TOBS in the region Randstad South (more specifically: in the area of district control office Rotterdam), will result in an expected FRT of 114.1 +28.0* 1 = 142.1 minutes (the other variables all get a value of zero).

These formulas can be used to predict the function recovery time based on the input variables discussed above. However, unfortunately this will only be a rough estimation. The adjusted R-squared value, delivered by the regression analysis, shows how well a model explains the data points. It is thus an indicator of the variability explanation of the model. In this case, the adjusted R-squared values for the models in Table 4.9 lie around fifteen percent per failure form. Thus, only fifteen percent of the variability is explained by the factors in the model. This value is low, which is why the model should not be used as an accurate prediction, but only as an indication. The low value isn't that surprising since the most important factor influencing the variance of the failure recovery duration is the cause, which isn't known beforehand. The failure forms with the highest adjusted r-squared values are the failure forms: level crossing barrier not responding properly and the failure form: path subsidence. Thus, for these two, the model should perform best during the theory testing phase.

The information of Table 4.9 can be used to, potentially, improve the confidence intervals determined earlier (see Table 4.5). The execution of this process is explained and executed in Appendix G. The resulting intervals per failure form and region are also presented in this appendix. The accurateness of these intervals will be tested during the theory testing phase.

Table 4.9: The Regression Equations predicting the FRT per Failure Forms based on the relevant Input Factors.

Failure Form	Regression Equation
Path TOBS (1)	FRT = 114.8 - 34.4*Grinding Train + 0*VL-Arnhem - 43.7 *VL-Amersfoort + 47.2 *VL-Alkmaar - 24.0* VL- Amsterdam - 7.8*Eindhoven - 71.9*VL-Groningen - 41.0*VL-The Hague + 49.4 VL-Kijfhoek - 32.9 VL*Maastricht + 67.3*VL-Roosendaal+ 28.0*VL-Rotterdam + 45.8*VL-Utrecht - 32.2 VL-Zwolle
Switch NIC (2)	FRT = 85.6 - 20.7*PGO-Contract
Level crossing disturbed (3)	FRT = 81.3 - 23.0*Rush-hour - 22.3*PGO-contract+ 0*VL-Arnhem - 9.6 *VL-Amersfoort- 41.0 *VL-Alkmaar + 5.3 *VL-Amsterdam - 5.8 *VL-Eindhoven - 26.9 *VL-Groningen - 44.5 *VL-The Hague - 29.6 *VL-Kijfhoek - 33.3 *VL-Maastricht - 16.7 *VL-Roosendaal - 29.7 *VL-Rotterdam - 21.3 *VL-Utrecht + 16.8 *VL-Zwolle
Level crossing collision (4)	FRT = 107.8 - 23.1*Rush-hour + 0 *VL-Arnhem + 19.6 *VL-Amersfoort + 22.2 *VL-Alkmaar + 14.4 *VL-Eindhoven - 17.5 *VL-Groningen + 2.4 *VL-The Hague - 5.3 *VL-Maastricht + 97.4 *VL-Roosendaal - 50.8 *VL-Rotterdam + 12.9 *VL-Utrecht + 4.5 *VL-Zwolle
Switch TOBS (5)	FRT = 121.8 - 20.6*PGO-contract
Sign our of order (6)	FRT = 87.7+ 0*VL-Arnhem - 51.1 *VL-Amersfoort - 18.3 *VL-Alkmaar - 22.7 *VL-Amsterdam- 61.7 *VL-Eindhoven - 41.0 *VL-Groningen - 54.2 *VL-The Hague + 63.1 *VL-Kijfhoek + 17.3 *VL-Maastricht - 38.1 *VL-Roosendaal - 70.7 *VL-Rotterdam + 64.9 *VL-Utrecht - 24.4 *VL-Zwolle
Level crossing barrier not responding properly (7)	FRT = 40.35 + 284.6*Maintenance Activities
Other failures (8)	FRT = 140.4 - 78.7*Rush-hour
Path subsidence (9)	FRT = 658.0 + 0*VL-Arnhem - 563 *VL-Amersfoort - 535 *VL-Alkmaar - 558 *VL-Eindhoven - 608 *VL-Groningen - 616 *VL-The Hague - 512 *VL-Kijfhoek - 596 *VL-Maastricht - 526.1 *VL-Roosendaal - 548 *VL-Rotterdam - 561.8 *VL-Utrecht - 597 *VL-Zwolle
Status path. Others (10)	FRT = 66.3 - 28.5*PGO-contract+ 95.3*Maintenance Activities
Switch disturbed (11)	FRT = 85.6 - 20.7*PGO-Contract
Level crossing specific damage (12)	FRT = 81.3 - 23.0*Rush-hour - 22.3*PGO-contract+ 0*VL-Arnhem - 9.6 *VL-Amersfoort- 41.0 *VL-Alkmaar + 5.3 *VL-Amsterdam - 5.8 *VL-Eindhoven - 26.9 *VL-Groningen- 44.5 *VL-The Hague - 29.6 *VL-Kijfhoek - 33.3 *VL-Maastricht - 16.7 *VL-Roosendaal- 29.7 *VL-Rotterdam - 21.3 *VL-Utrecht + 16.8 *VL-Zwolle- 23.0*Rush-hour - 22.3*PGO-contract
Buckling in path (13)	FRT = 66.3 - 28.5*PGO-contract+ 95.3*Maintenance Activities
Sign turned off (14)	FRT = 87.7+ 0*VL-Arnhem - 51.1 *VL-Amersfoort - 18.3 *VL-Alkmaar - 22.7 *VL-Amsterdam- 61.7 *VL-Eindhoven - 41.0 *VL-Groningen - 54.2 *VL-The Hague + 63.1 *VL-Kijfhoek + 17.3 *VL-Maastricht - 38.1 *VL-Roosendaal - 70.7 *VL-Rotterdam + 64.9 *VL-Utrecht - 24.4 *VL-Zwolle
Smouldering sleepers (15)	FRT = 66.3 - 28.5*PGO-contract+ 95.3*Maintenance Activities
Level crossing TOBS (16)	FRT = 81.3 - 23.0*Rush-hour - 22.3*PGO-contract+ 0*VL-Arnhem - 9.6 *VL-Amersfoort- 41.0 *VL-Alkmaar + 5.3 *VL-Amsterdam - 5.8 *VL-Eindhoven - 26.9 *VL-Groningen- 44.5 *VL-The Hague - 29.6 *VL-Kijfhoek - 33.3 *VL-Maastricht - 16.7 *VL-Roosendaal- 29.7 *VL-Rotterdam - 21.3 *VL-Utrecht + 16.8 *VL-Zwolle- 23.0*Rush-hour - 22.3*PGO-contract
Bridge disturbed (17)	FRT = 66.3 - 28.5*PGO-contract+ 95.3*Maintenance Activities
Overpass collision (18)	FRT = 66.3 - 28.5*PGO-contract+ 95.3*Maintenance Activities

The equations given in Table 4.9 are obviously a bit overwhelming and impractical to work with. Since the accuracy is also relatively low (due to the low r-squared value), a more practical, general summary of the performance of the district control offices is given. This is done based on the coefficient every district control office has per failure form. By comparing this to the average coefficient a district control office has on that failure form, the performance of that specific district control office per failure form is determined. By averaging the performance of a post over all the failure forms , the general performance of a post is calculated. This divides the posts in five classes, ranging from a good performance to bad performance compared to the average:

- 30 Minutes faster than average, district control offices: The Hague and Groningen.
- 10 Minutes faster than average, district control office: Amersfoort, Maastricht, Rotterdam and Zwolle.
- Average, district control offices: Amsterdam and Eindhoven.
- 10 Minutes slower than average, district control office: Arnhem and Alkmaar.
- 30 Minutes slower than average, district control offices: Kijfhoek, Roosendaal and Utrecht.

These performances thus only apply on the failure forms where the district control office is a variable of influence (thus not the switch failure forms).

Nearest Neighbour

Nearest neighbour is a common pattern recognition method used as a theory building method now. During the method the response value (in this case: the FRT) is predicted by comparing the situation to previously occurred situations and their corresponding response values.

The method consists out of two steps: neighbour recognition and the prediction of the response value. According to literature: during the neighbour recognition step the situation is compared to a set of previously occurred data points based on certain variables of the situation. When a data point is, based on these variables, close enough to this particular occurrence it is called a neighbour. In the second step, the response values of the neighbouring values are used to predict the response value of the actual data point (Dasarathy, 1991).

When this is applied on the FRT-data, the variables used during the regression analysis are used to determine neighbouring situations. These variables are: the failure form, rush-hour, the contract-type, recent work of the grinding train and recent maintenance on the railroad track. The region the failure occurs in is also used as a variable. The choice to use the region as a variable, as opposed to the more specific district control office, is because of the relevance of the other variables and the high amount of district control offices. It is important that, when determining neighbouring situations, these variables have the same value but the amount of possible situations is not too high. However, an exact match of situation will quickly result in a high amount of possible situations: 18 (failure forms) * 2 (rush-hour)*2 (contact-type) *2 (grinding train) *2 (maintenance) = 288. Adding the variable district control office with thirteen options will increase the amount of occurring situations to an unmanageable amount. While, using the four different region types with 4*288 = 1152 possible situations is slightly better.

In practice, this method results in 275 different situations failures occurred in last year. 216 Of these failures occurred more than once and therefore have a prediction based on previous data. Occurrence rates vary widely: 59 situations thus only occurred once over a year, while 877 did not occur at all. During the same year, two situations occurred more than a 100 times (a path TOBS and level crossing disturbed in the region North East while the other variables all are zero).

Most of the situations however, occurred relatively little. Therefore, the prediction of the FRT for these situations is not as accurate as possible when more data would have been available. When the method is applied over multiple years this accurateness will increase automatically. In order to judge the performance of the nearest neighbour method only the most occurring situations are therefore judged during the theory testing phase in the next chapter. The situations which occur at least twenty times are selected. This results in a manageable number of reliable cases and in an accurate prediction.

The prediction of the FRT per situation with at least twenty occurrences in one year is shown in Table 4.10. The accompanying intervals are also giving. This allows the nearest neighbour method to be compared with the other methods in chapter 5. A type number is handled here to indicate the situation. The type number illustrates the value of the variables in the following order (left to right): failure form, region (1 is region South, 2 is region North East, 3 is regions Randstad South and 4 is regions Randstad North), rush-hour, contract-type, grinding train and recent maintenance. A value of one shows the activeness of a variable while zero shows the absence. Thus, type number 120100 illustrates the situation where a path TOBS occurs (see Table 4.9 for the link between failure number and failure form) in region North-East with a PGO-contract. There is no recent grinding train or maintenance activities and it does not occur during rush-hour. In order to predict the failure recovery time, the average of the previous occurrences is taken. For this failure situation the expected FRT is therefore 85.3 minutes.

Table 4.10: The FRT-Prediction per Type based on the Nearest Neighbour Method applied on the Failure Data from September 2014 to August 2015.

Type Number	FRT- Prediction (minutes)	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
110000	136	188	231	239	47.5%	57.6%	65.0%
110100	76	196	231	414	38.5%	53.8%	63.5%
120100	85	120	136	162	43.8%	61.2%	67.8%
121100	55	56	79	82	51.4%	74.3%	85.7%
130100	80	67	87	231	57.7%	76.0%	87.5%
140000	78	150	178	233	28.6%	50.0%	60.7%
141000	153	202	264	270	19.1%	35.0%	36.8%
210100	74	68	75	131	31.3%	64.5%	90.0%
220100	87	84	92	118	50.0%	70.2%	78.3%
221100	91	85	87	216	33.3%	60.0%	84.2%
230100	47	44	47	51	60.0%	88.0%	96.0%
240000	68	77	112	144	18.8%	52.1%	77.1%
310000	52	95	116	141	60.7%	71.4%	75.0%
310100	54	48	68	95	64.1%	79.5%	84.6%
320100	56	78	141	211	51.0%	67.6%	79.4%
321100	56	61	70	87	50.0%	75.0%	83.3%
420100	66	76	82	86	44.6%	66.1%	78.6%
520100	69	57	68	71	45.5%	77.3%	90.9%
620100	108	79	108	125	52.4%	57.1%	66.7%

4.3. Conclusion

During this chapter new failure forms are determined in order to decrease the variation of the failure recovery time (FRT). The methods determined in chapter three are executed in order to form predictions for the FRT. Based on this theory building step, the models are tested in the next chapter in order to determine the best prediction model and to form the predictions per failure form self. The temporary failure recovery actions are analysed in Appendix I.

The first part of this chapter shows that the distinction between twenty-two failure forms can be made. However, the four lowest occurring failure forms are omitted from this research due to their low occurrence (less than one percent of total failures), leaving eighteen failure forms. The four most occurring failure forms are: path TOBS, switch NIC, level crossing disturbed and level crossing collision. These represent more than 60 percent of the total amount of failures.

For each of these eighteen failure forms an initial prognosis has to be determined. In order to work with the data from the SAP-system the waiting time of the mechanic for a path withdrawal had to be deleted from the data. This is done in section 4.1, leaving the diagnosis time and the repair time suitable for analysis during the remainder of the chapter.

The first method applied, prediction by the mechanic, confirmed the difficulty of the task. The mechanics could not accurately predict the length of a failure recovery based on the failure form only, since this is too dependent upon the failure cause. They can, however, judge the FRT of failure forms relatively to each other and give practical reasons for these lengths. The ranking of the FRT per failure form can be seen in Figure 4.5 and ranges from a sign failure to a path subsidence.

Determining the FRT by its probability distribution delivered more practical solution in the form of percentiles and percentage of failures recovered after certain time periods. This method showed that the FRT follows a different probability distribution per failure form ranging from 3-parameter Weibull to exponential and normal distributions. This also resulted in extremely different percentiles and failure recovery percentages per failure form. The sign- and level crossing-failures, for instance, can be repaired in the first 25-minute period 60 percent of the time while this is only 40 percent for a path TOBS. It also showed the large difference between the repair times of a regular and a diamond switch (see Figures 4.6-4.7).

The other methods: regression analysis and nearest neighbour introduced more variables besides the failure form. Combined with the intervals mentioned earlier, this results in intervals based on more specific situations for the regression analysis method. This shows that a path TOBS takes approximately 60 minutes longer to be repaired in the areas of district control offices: Kijfhoek, Roosendaal and Utrecht than in the area of district control offices Groningen and The Hague. Another conclusion found by this analysis is the influence of the PGO-contract on the FRT of a switch NIC-failure. In general, this failure form is repaired twenty minutes faster in an area with this type of contract in comparison with an area with an OPC-contract. The influence of the grinding train on the failure form: path TOBS is also significant: the failure recovery time is, on average, reduced by 40 minutes when a grinding train has recently passed.

These models are only formed by the prediction methods during the theory building phase. In the next chapter, the models are tested on new failure data in order to determine how they perform. The best model can hereby be selected. Besides this, practical issues can be found in this step.

5. Theory Testing

In chapter four the methods found in chapter three are applied on the data of ProRail. This results in multiple models displaying the predicted failure recovery times of different failure forms. In this chapter, the models built in the last chapter are tested. A new set of data is taken and, by applying the models, the FRT is predicted. This enables us to score each model based on its prediction performance. This results in the best possible FRT-prediction model.

5.1. Data Set

The models developed in the last chapter are based on the data set from September 2014 till August 2015. To test the models developed, a new data set is used. Therefore, the latest year of data not already used is selected. This data set consists out of two periods: December 2013-August 2014 and September 2015 – November 2015. By selecting the data from an entire year, possible seasonal effects are eliminated. Since this data is recent, the models developed are robust enough to withstand recent improvements in failure recovery. Again, the data is selected by the steps displayed in Table 2.1 but now applied on the new data set. Thus, the failures used are of priority levels one and two and only TAO-failures of failure recovery type one. Just like previous data sets, the data is cleaned by removing missing, noisy and inconsistent data points. A problem with this database is the low amount of data from the 2013 data set. During this year, data monitoring was not done as strictly as during 2014 and 2015 (for instance: the failure form was not stated beforehand), thereby resulting in way less useful data than the other years. The implications of this are explained later. The new data set is from now on referred to as: theory data set.

The resulting data points of the theory data set are classified into their failure forms again (see appendix D for the explanation of every failure form). During this chapter, only the six most occurring failure forms are evaluated (path TOBS till sign out of order). These six failure forms represent more than 70 percent of all the failures. Due to their higher occurrence the accurateness of their prediction is higher than the other failure forms and thus more reliable. Besides this, the low amount of these failure forms enables us to look more specifically at the disturbances in the prediction and fix these.

5.2. Performance of the Methods

The models developed in the last chapter are evaluated one by one starting with the prediction of the FRT by the expert.

Prediction Expert

The failure recovery experts (mechanic) cannot estimate the exact duration of a failure recovery for a specific failure form. However, they are capable of scoring the failure forms in comparison to each other (see Figure 4.5). When only the top six failure forms are taken into account, this is in the following order (from a low FRT to a high):

- Switch TOBS;
- Switch NIC, level crossing disturbed & sign out of order;
- Path TOBS & level crossing collision.

Since the expert based this comparison on the expected FRT, this model can be compared by looking at the average FRT of the failure forms from the theory data set. These times are shown in Table 5.1.

Table 5.1: The Average FRT of the Six most occurring Failure Forms during December 2013 till August 2014 and September 2015 till November 2015.

Failure Form	Average FRT
Switch TOBS	50.9
Switch NIC	46.3
Level Crossing Disturbed	31.3
Sign Out of Order	35.5
Path TOBS	55.8
Level Crossing Collision	76.2

From the table it is concluded that the experts did not predict the FRT completely correct. They underestimate the duration of a TOBS in a switch section while the TOBS in a path section is not as bad as expected. However, the similar FRT of an irregular sign failure and a disturbed level crossing are correctly predicted. The experts also correctly predicted the long FRT for a level crossing collision.

Confidence Intervals

In chapter 4 we determine the probability distribution per failure form. Via this distribution the 50-, 80-, 85- and 90- percentiles are determined. Besides this, the amount of failures solved within one, two and three 25-minute blocks became clear. For the theory data set these same boundary values can be determined and compared to the already developed boundary values. The boundary values of the new data set are taken by looking at the corresponding value in the data set. For instance: for the 50-percentile, the 100th value of a data set containing 200 values is taken (50%*200 = 100). For the 80-percentile, the 160th (80%*200) value is taken. This is a quicker method than via the probability distribution, but, since the distribution is already clear, suitable to check the performance of the distribution. Table 5.2 shows the difference between the percentiles and the percentage of data points in the 25-minute intervals of the probability distributions (see Table 4.5) and the percentiles of the theory data set. For the percentiles, a positive value indicates an underestimation of the previously determined percentile while a negative value indicates an overestimation. Thus, for the theory data set, the 50 percent repair value of a path TOBS is 10.9 minutes lower than the interval value. For the 25-minute blocks (the largest possible gap between train traffic), a positive value indicates that the theory data set contains more data points within this block than expected while a negative value is the opposite. Thus, the first 25-minute block for a path TOBS contains, in the theory data set, 1.5 percent more of the data points than expected (41.1 percent in comparison with the expected 39.6 percent) while the first block for a switch NIC - diamond contains 0.7 percent less data points than expected.

Table 5.2: The Difference between the Determined Percentiles (in Minutes) and the Interval Values of the Theory Data Set.

	50% Repaired	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Path TOBS	-10.9	-27.0	-37.1	-36.2	1.5%	5.9%	7.8%
Switch NIC - Regular	-3.7	-3.9	-3.8	0.4	0.8%	0.2%	3.0%
Switch NIC - Diamond	-5.3	-2.2	-11.0	-5.1	-0.7%	-3.2%	-6.3%
Level Crossing Disturbed	5.2	-4.2	-5.2	3.2	-7.1%	2.0%	7.0%
Level Crossing Collision	25.3	14.7	12.3	3.1	-23.4%	-26.6%	-10.8%
Switch TOBS	-5.9	-5.0	-10.9	-11.8	-0.1%	-1.5%	-0.7%
Sign Out of Order	-3.8	-11.0	-13.0	-14.8	1.7%	4.1%	4.1%

As can be seen from the table, some of the failure forms perform relatively well while others do not. It is important to note the unreliability of the values in the 90- percentile. Due to the relatively low amount of data, picking a value on the far end of the distribution (the 90 percent value) can easily result in a big over- or under-estimation. This performance indicator is therefore not reliable. The most reliable performance indicators are the 50- and 80- percentiles and the 50 minute interval because they represent the less extreme data points.

As can be seen, the path TOBS and the level crossing collision failure forms don't perform well. A visual check of both the data sets also confirms the quite large differences between the two data sets. This is illustrated in Figure 5.1. This figure shows the I-MR chart of the failure form: Path TOBS created with the help of the software program Minitab (a program specialized in analysing production processes). In the upper section the individual values of failure form are shown in the order they occurred. The lower section shows this as well but now for the variation (based on the moving range[6]). Each of the sections is accompanied with an upper control limit (UCL) and a lower control limit (LCL) which represent a distance of three standard deviations from the average.

The I-MR Chart is a control tool which indicates when a certain value or a sequence of values behaves differently than expected. When this happens, Minitab places a red square with its accompanying error code at the corresponding value(s). These are also shown in figure 5.1. The error codes have different meanings, for instance: error code 1 indicates a value above the UCL, error code 2 shows nine values on the same side of the centre line and error code 5 indicates

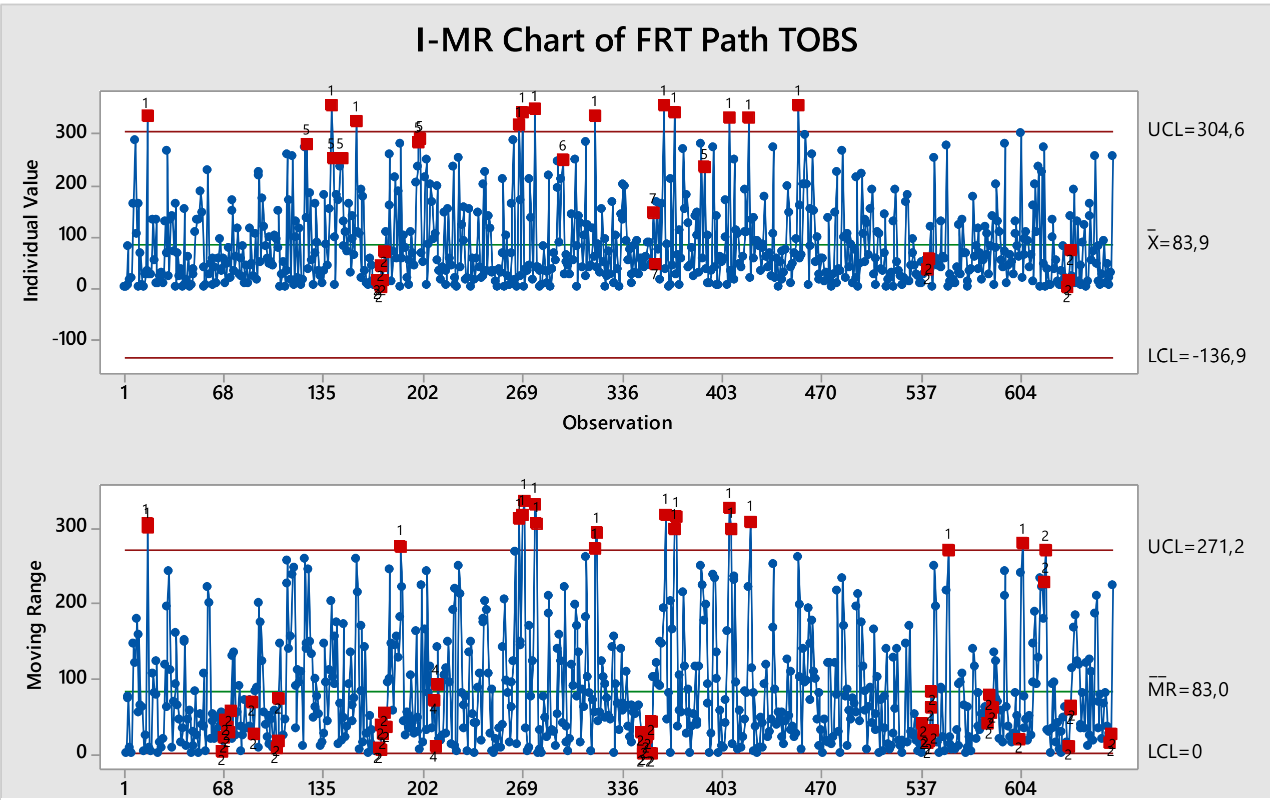


Figure 5.1: The I-MR Chart of the Failure Form: Path TOBS for data from September 2013 till November 2015. The Upper Section shows the Individual Values while the Lower Section shows the Variation via its Moving Range.

[6] The moving range is the difference between one observation and the observation immediately prior to this one (Quality America Inc., 2016).

that two out of three values are at a distance of more than two standard deviations from the centre line.

Important here is the difference between the two data sets. The original data set (September 2014 till August 2015) ranges from data point 70 till 480. The theory data set contains the other data points. As can be seen from Figure 5.1, more type one errors occur during the period of the original data set. However, this can be explained due to the fact that this data set was, on average, higher and therefore contained more values above the upper limit. Besides this, no big differences can be spotted: the usual amount of low data points are present, the amount of high values (not the really high) is the same and the variance behaves the same (also with the occasional type 2 errors).

The same figure can be created for the failure form: level crossing collision. This chart is shown in figure 5.2.

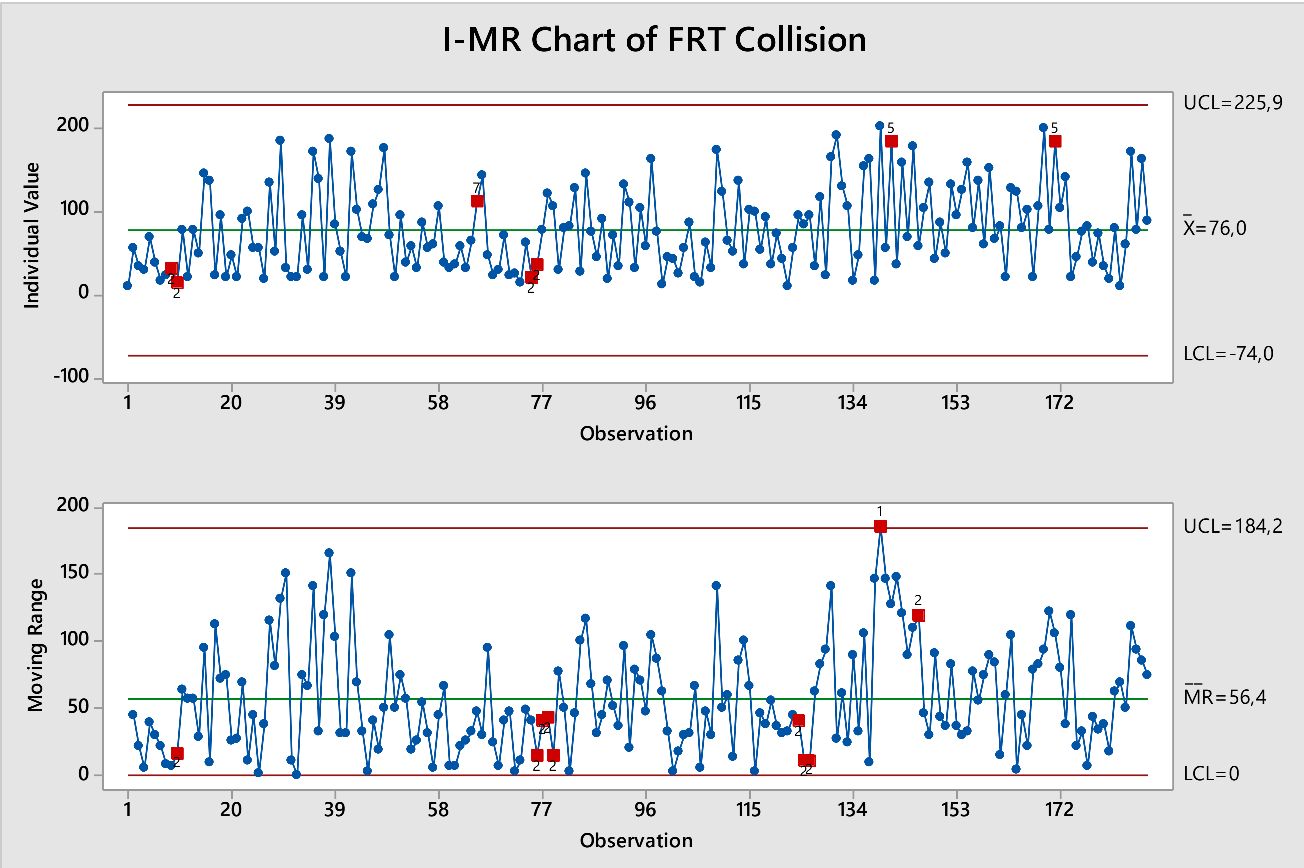


Figure 5.2: The I-MR Chart of the Failure Form: Level Crossing Collision for the Data from September 2013 till November 2015. The Upper Section shows the Individual Values while the Lower Section shows the Variation via its Moving Range.

The original data set for the failure form: level crossing collision starts at data point 15 and ends at data point 145. The other data points are from the theory data set. Again, no big differences in the behaviour of the two data sets can be spotted. It can thus be concluded that the percentile differences between the data sets are not due to a change in behaviour of the FRT.

The average FRT of a path TOBS is higher in the original data set while the FRT of a level crossing collision is lower in this data set. Since the behaviour didn't change, there tend to be differences between the years of these failure forms. In order to make the percentiles more robust for these differences, the values for the failure forms: path TOBS and level crossing collision are averaged out over these years based on Table 5.2. For instance: the 50– percentile for a path TOBS is lowered by 5.5 minutes (10.9/2). The new intervals are shown in Table 5.3.

Table 5.3: The Improved Intervals (in Minutes) for the Failure Forms: Path TOBS and Level Crossing Collision based on Data from 2013, 2014 and 2015.

	50% Repaired	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Path TOBS	39.7	109.1	129.4	165.7	40.3%	57.5%	69.6%
Level Crossing Collision	52.9	105.3	122.3	142.8	29.5%	45.4%	66.1%

Regression Analysis

For the regression analysis the possible regression variables are analysed one by one, starting with the different district control office. A distinction between them is made based on the five classes as identified at the end of the regression analysis chapter. These range from best performing class 1: district control offices Groningen and The Hague, to worst performing class 5: district control offices Kijfhoek, Roosendaal and Utrecht. To compare these classes, the average of their FRTs per failure form for the theory data set is taken which can be compared to the expected difference between classes as shown in chapter 4. These are shown in Table 5.4. Since the factor district control office doesn't influence the switch failure forms, these are not taken into account.

Table 5.4: The Average FRT (in Minutes) per Failure Form and district control office Class for the Theory Data Set.

	Class 1	Class 2	Class 3	Class 4	Class 5
Path TOBS	42.2	51.2	62.2	68.7	71.7
Switch NIC	-	-	-	-	-
Level Crossing Disturbed	47.3	30.4	53.1	56.3	38.6
Level Crossing Collision	61.0	89.3	93.4	70.4	112.0
Switch TOBS	-	-	-	-	-
Sign Out of Order	38.5	45.6	49.3	20.4	75.6

Again, the differences between the district control offices can be seen. These are strongest present in the failure form: Path TOBS. The differences, however, are not as big as predicted (a full hour difference between the average of class 1 and class 5) but about half as strong.

The influence of the PGO-contract is looked at next. Therefore, the average FRT of failures of the failure form: switch NIC with and without an PGO-contract are compared.

- Average FRT with PGO contract: 51.8 minutes.
- Average FRT without a PGO contract: 67.4 minutes.

Thus, in general the percentiles of the FRT of a failure with a PGO-contract are eight minutes lower than average while the percentiles of the FRT of a failure without an PGO-contract is eight minutes higher than average.

With the same method, the effect of the grinding train is analysed. This shows a lowering of the FRT of a path TOBS with thirty minutes when the grinding train has recently passed: like expected. Since the amount of data influenced by this train is low, the normal situation, without the grinding train, does not have to be adjusted like for the PGO-contract.

The FRT of a repair during rush-hour does not show any difference with the FRT of a repair outside of rush-hour for the failure forms of a level crossing. This is not surprising considering the high test-statistic of the variable rush-hour (see Table 4.7) and therefore the low reliability. This effect is therefore removed from the research.

The last regression variable to be tested is the execution of recent maintenance activities. However, after several conversations with VL (traffic control) it is decided to remove this effect from the percentiles. Not because the effect is not significant, but because the effect can't be generalised with a number. According to VL, the type of maintenance executed, the location where this is done and the expected mistakes made, all influence the FRT advantage (or disadvantage). A general number is therefore not robust enough to illustrate the effect of this variable while the VL is capable of determining this relatively easily by consulting the contractor about the recent maintenance activities. It is therefore concluded that recent maintenance activities at the failure location have a significant effect on the failure recovery time and should therefore be considered, but the specific impact should be judged by VL themselves.

Nearest Neighbour

The nearest neighbour model predicted the FRT by looking at previous FRT-values of failures occurring in the same situation. Based on this, Table 4.9 shows the predicted value, and the intervals from the twenty most occurring situations. When this is applied on the theory data set it becomes clear that an accurate test of the intervals is not possible due to the low amount of data per situation. The comparison of the prediction and the average from the theory data set can be done and is shown in Table 5.5. This table also shows the difference between the two.

This difference is pretty large in most situations: in eight out of the twenty situations more than half an hour. Some of these differences are due to the low amount of data and the lack of good predictors (low r-squared value) but it still underperforms. Since even the most occurring situation, which an exact match to other occurrences, does not give a good prediction, generalizing or gaining more observations will not give a suitable solution. We therefore conclude that the nearest neighbour method is not a reliable model to predict the FRT.

Table 5.5: The Prediction by the Nearest Neighbour Method, the Actual Values and the Difference between the Two for the Twenty most Occurring Situations.

Type Number	FRT- Prediction (Minutes)	Theory Data Set Average (Minutes)	Difference (Minutes)
110000	135.5	59.6	75.9
110100	76.2	77.7	-1.5
120100	85.3	58.0	27.3
121100	55.1	65.3	-10.2
130100	79.6	98.3	-18.7
140000	78.3	98.9	-20.6
141000	152.6	63.0	89.6
210100	74.1	55.3	18.8
220100	87.3	45.6	41.7
221100	91	38.2	52.8
230100	47.1	69.0	-21.9
240000	68.0	44.0	24.0
310000	51.5	20.6	30.9
310100	54.1	69.4	-15.3
320100	56.3	42.8	13.5
321100	55.8	45.4	10.4
420100	66.2	84.8	-18.6
520100	68.5	26.0	42.5
620100	107.7	46.5	61.2

5.3. Conclusion

This chapter introduced a new data set: December 2013- August 2014 and September 2015–November 2015. Based on this data set the performance of the models developed in chapter four is determined and the best model/combination of models selected. It can be concluded that the expert can relatively well predict the easiest and hardest failure forms to repair. However, they tend to overestimate the duration of a path TOBS and underestimate the duration of a switch TOBS.

By comparing the theory data set with the percentiles we conclude that these percentiles perform relatively well apart from two, frequently occurring, failure forms: path TOBS and level crossing collision. There are large differences between the two data sets. By adjusting the percentiles of these two failure forms to the percentiles of Table 5.3, the FRT of these failure forms is made more robust for fluctuations over the years. The percentiles for the other failure forms, especially the switch failures, only show minor differences with the theory data. We therefore conclude that these percentiles are reliable.

The influence of the district control offices again showed to make a difference in the new data set, however, to a lower extend. To make the intervals robust over more years the results are combined and the following differences between the five classes of district control offices are concluded:

- 20 Minutes faster than average, district control offices: The Hague and Groningen.
- 10 Minutes faster than average, district control offices: Amersfoort, Maastricht, Rotterdam and Zwolle.
- Average, district control offices: Amsterdam and Eindhoven.
- 10 Minutes slower than average, district control offices: Arnhem and Alkmaar.
- 20 Minutes slower than average, district control offices: Kijfhoek, Roosendaal and Utrecht.

The other factors: contract type, rush-hour, grinding train and maintenance activities showed fluctuating results. The use of a PGO-contract again showed an improvement of ten minutes of the percentiles of switch failures and the recent use of the grinding train lowered the expected FRT of a path TOBS with thirty minutes. However, the effect of a repair during rush-hour proved non-existent while recent maintenance activities are, after consultation with VL, too dependent upon the exact situation to be generalizable. To conclude: the recent passing of a grinding train, a PGO contract and recent maintenance activities at the failure location all influence the expected FRT. While the first two have a generalizable effect, the latter depends too much upon the situation and should therefore be determined by the traffic control (with the help of the mechanic/contractor) when the situation appears.

The nearest neighbour model did not provide a reliable prediction model. More data might improve this, but for now, this model can't be used to predict the FRT.

6. Detailed Look at the Underlying District Control Office Variables Influencing the Failure Recovery Time

One of the identified variables influencing the failure recovery time (FRT) significantly is the district control office. However, the general perception of the traffic control (VL) is that there have to be variables in relation with them automatically accompanying certain district control offices. By revealing these variables it is possible to actively work on decreasing their negative impact and learn how things can be improved. Besides this, it is also possible to determine intervals suitable for more specific failures instead of generalizing them over the entire district control office region. Due to the fact that most of the suspected causes are variables determined for the first time, they are not in the SAP-database. Thus, another method to reveal these is necessary. Therefore, this is treated in this separate chapter. First, the method which is used will be explained. Next, the possible variables researched in this chapter are explained. This is followed by the identified impact of the variables and the impact of this on the intervals and an example.

6.1. Methodology

In order to acquire the new variables which explain the differences between the district control offices and enable a more specific FRT prediction, another method as before has to be applied. This is due to the fact that most of the variables researched are not present in the databases or actively registered.

1. *Determine possible explaining variables.*

A list of variables/theories which may explain the differences between the district control offices is determined first. This is done by interviewing experts about the topic. This includes, for example, infrastructure-experts and traffic control-employees.

2. *Research variable per district control office.*

When a list of possible variables is made, a number of district control offices have to be visited in order to research the variables per district control office. Visiting the posts is necessary due to the fact that not all the information is available in the databases. It is important to visit district control offices which have different FRT-performances. By doing this, the variables responsible for the differences are easier to spot. Per variable, the results of this research will be different: it can be a performance score from zero to ten or a specification of the route sections ("baanvakken") where this variable is active and where not. The effects of this are explained in the next step.

3. *Compare variable results with district control office performance.*

The results per variable per district control office have next to be compared with the failure recovery time performance of each district control office. With the help of, for instance, regression analysis and t-tests, it is possible to determine which of the variables on the list have a significant effect on the failure recovery time and which don't. Besides this, the quantitative effect can be determined. The method to do this differs per variable because the results per variable will have different formats, for instance: a t-test can be used to determine a difference when a variable is active or not, regression analysis is more suitable when a score from one to ten is given for the variables.

4. *Update percentiles and intervals.*

When the quantitative effect of every variable is clear, the link with the previously determined percentiles and intervals can be made. These have to be updated based on the new information. It is also important to keep the information manageable. Thus, to use a tool to make it possible to take every variable into account when the percentiles and intervals are determined.

6.2. District Control Office Specific Variables

Based on interviews with infrastructure-experts (of the ProRail department: Architecture and Technique) and VL-employees a list of variables possibly explaining the differences between district control offices is determined. The variables and the reason why they cause differences are as following:

- *Sharp Corner in the route section ("baanvak").* A baanvak is a part of the path between two stations (or, occasionally, other significant points influencing the train schedule like bridges and maintenance points). A sharp corner increases the amount of wear of the track on that part of the path. This causes more metal particles to appear and more TOBS's caused by these.
- *Age of the Train Rails,* how long ago the rails ("spoorstaven") were replaced. An older rails increases the likelihood of metal particles occurring and being the cause of a TOBS.
- *Movability around the Path*, the possibility for the mechanic to get easy and quick access to every part of the path relevant for repairing the failure, thus: how easy he can move around the path. This is relevant for the failure form path TOBS where, occasionally, large distances have to be traveled either by foot or by car, depending on the movability around the path.
- *Contractor Selection Continuity,* the same contractor has been used for the last years. A new contractor usually does not know the failure history of the infra-equipment present in that area or the fastest access methods to parts of the infrastructure. This possibly increases the failure recovery time.
- *Distance of Spare Parts,* the distance a certain route section is from the depot. A high distance increases the repair time when a spare part is necessary. Whether or not a spare part can be picked up from the local depot (the depot of a contractor in a specific district control office area) or has to be picked up from the central depot far away also influences the FRT.
- *PCA,* the contractor(s) operating in the area of the district control office. Differences between the performances of contractors per failure form may exist.

Whether or not a variable explains the different FRT and its exact effect are researched in section 6.3.

6.3. Influence of Variables

For this section, multiple district control offices were visited or contacted (Alkmaar, Groningen, Maastricht, The Hague, Utrecht and Zwolle) to check their scores on the variables explained above. This results either in a score for the entire region or specific route sections or in an expected effect of the variable. The research conducted per variable is given in Appendix H.

Appendix H identified that most of the variables of section 6.2 have a significant influence on the FRT. However, the distance of the spare parts follows from the PCA active in that district control office, it is thus already included in another variable and is excluded from this research. The variable contractor selection continuity has a significant effect on the FRT, but, since the contracts only possibly swap once in five years and the limited available data, it is ignored in this research.

The variables give results (thus explained in appendix H) which have to be used in different ways, variable PCA being the most difficult. Table 6.1 gives the result of the variable: PCA. Per failure form the PCA(s) with the lowest expected failure recovery time (FRT) is (are) given by a number or a dash. The dash means that there is no significant difference between the PCA performing best and this PCA. When a PCA performs worse, this is shown by the number with the plus symbol in front of it. This shows the average amount of minutes this PCA takes longer to repair a specific failure form.

For example, Asset-Rail, Strukton and BAM Rail both need 50 minutes, on average, for a level crossing collision while the FRT of Volker Rail takes 34 minutes longer.

Table 6.1: The Average FRT (Minutes) of the Best Performing PCA and the Difference of the other PCAs per Failure Form.

	Asset-Rail/Strukton	BAM Rail	Volker Rail
TOBS Path	56	+17	+34
Switch NIC	35	-	+27
Level Crossing Failure	+5	27	+12
Level Crossing Collision	50	-	+34
TOBS Switch	36	-	-
Sign Out of Order	+14	24	+22
Status Path	+22	+5	55

Besides the PCA, the remaining three factors all had a significant effect on the FRT as well.

When the accessibility and possibility to move around the path is good, the expected FRT of a TOBS decreases with ten minutes, while a bad movability increases this with twelve minutes. A sharp corner in the route section ("baanvak") followed by a TOBS increases the likelihood of metal particles being the cause of the failure and therefore lowers the expected FRT with thirty minutes. The age of the train rails also influences the FRT of a TOBS, again by metal particles causes by extra wear of the rails. In this case, the effect isn't as significant as a sharp corner but still relatively large. A very old rails with a TOBS usually is recovered fourteen minutes faster than a normal rails. Keep in mind that this research is about the FRT and not about the expected amount of failures. An old rails may be easier to repair but it is not superior to a new rails because the extra wear and metal particles will cause a higher failure rate.

When applying the variables on the expected FRT it is important to note that from the variables increasing the odds for metal particles (grinding train, old rails and sharp corner) only the highest effect of the variables active has to be applied. This prevents that the probability of metal particles is applied more than once.

6.4. Practical Application

With all the failure forms, different path sections and different variables influencing the failure recovery time (FRT) it is important that a structured method is used to arrive at the percentile suitable for the situation. Therefore, the following steps have to be used:

1. Select the percentile from Table 4.5 or Table 5.3 (use the latter in case of a path TOBS or a level crossing collision) and the probability distribution from Table 4.3. When analyzing a temporary failure recovery, use Appendix I.
2. Based on the failed infra-equipment, determine:
    a. In case of a TOBS, whether or not the failed section is an assenteller- or a PSSSL-section (see appendix F).
    b. In case of a TOBS, whether or not the equipment is very old.
    c. In case of a switch failure, whether or not the contract type is PGO
3. Based on the failure location, determine:
    a. The PCA active at that location
    b. In case of a path TOBS, whether the movability around the failure location is: difficult, medium or easy.
    c. In case of a TOBS, whether or not a grinding train recently passed the location
    d. In case of a TOBS, whether or not the failure is located near a sharp corner.
4. Combine the variables and the probability distributions to form the final percentiles.

<u>Example:</u>

In Zwolle a path TOBS occurs in an area with difficult movability around the path to a regular section (thus: no assenteller- or PSSSL-section). The other variables are not present.

Based on step 1, the percentiles for a path TOBS from Table 5.3 are selected. These are shown in Table 6.2.

Table 6.2: The Percentiles (Minutes) and Probabilities of Succeeding in Repairing the Failure in a Certain Block for a Path TOBS.

	50% Repaired	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Path TOBS	40	109	129	166	32%	58%	70%

Furthermore, the path TOBS has an exponential distribution (see Table 4.3) with an average of 88.31 and thus, a lambda of 1/88.31 = 0.0113.

While nothing has to be done based on step 2, step 3 has some implications. The PCA active in that area is Volker Rail, thus the average FRT is 34 minutes higher. The movability around the infrastructure is medium, thus not influencing the expected time. The average FRT is thus increased by 34 minutes.

During step 4, this increase in average has to be translated to the percentiles via the probability distribution. Therefore, the two factors (the previous probability distribution and the effect of the variables) can be seen as two separate distributions which have to be summed. Since the expectation of the sum of two random variables is the sum of the expectations (E[X+Y] = E[X] +E[Y]), the expected value of the new distribution is 88.31 + 34 = 122.31. The assumption is made that the distribution type stays the same (exponential), making the new lambda: 1/22.31 = 0.00818. The resulting percentiles are shown in Table 6.3.

Table 6.3: The Percentiles (Minutes) and Probabilities of Succeeding in Repairing the Failure in a Certain Block for a Path TOBS after applying the Variables.

	50% Repaired	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Path TOBS	51	138	166	205	37%	51%	62%

As can be seen from the percentiles, it takes a long time to repair this failure. It is therefore not advised to do this during the day. It should be kept in mind, though, that this is not a favourable situation.

In this case, with a path TOBS, the FRT behaved according to an exponential distribution. This is, however, not always the case. For the other distributions, accompanying other failure forms, the scale parameter has to be adjusted in the same fashion as shown in the example with the exponential distribution. Since the scale factor is the main factor influencing the outcome, this is a suitable method without complicating and lengthening the operation unnecessarily.

In addition to this report, a tool was developed which automatically determines the percentiles based on the failure form (and thus, probability distribution) and the values of the variables.

6.5. Model Validation

In the previous section a practical example of the model in action is shown. This example determines the FRT-intervals accompanying a path TOBS in area Zwolle repaired by the contractor Volker Rail while the movability is medium. This section evaluates how well these intervals but, in particular, the model is performing when compared with data from the actual situation.

Since the model is a combination of different prediction methods, there is not a standard evaluation way (like the R-squared value for a regression analysis). This is further complicated by the low amount of data for most cases due to the specific situations required. Therefore, we only analyse five cases which occur frequently and thus, have sufficient data to evaluate the intervals. These are the following five cases:

- *Case 1:* Path TOBS in Zwolle, Volker Rail, medium movability.
- *Case 2:* Path TOBS in Groningen, Strukton, medium movability.
- *Case 3:* Switch NIC in The Hague, Bam Rail, PGO-contract.
- *Case 4:* Switch NIC in Utrecht, Volker Rail, no PGO-contract
- *Case 5:* Switch NIC in Eindhoven, Strukton, PGO-contract.

In Table 6.4, we compare the intervals as predicted by the model with the actual situation per case. The difference between the two is also given. At the bottom of the table, the average of the differences is given per interval type.

The table shows different performances between different cases. Especially the second case: a path TOBS in Groningen performed differently than predicted by the model. This can be explained by the low amount of data (25/30 data points), a few unlucky data points will significantly increase the deviation between the model and the reality. Keeping this in mind, the model performances relatively well. The 25-minute blocks are, on average, five percent off while the 50-percent repair border is usually correct within 5/10 minutes. The 90-percent border shows a relatively large average difference, however, this can be explained by fact that this border is formed by the outliers: the high values. A 24 minute difference is only 15 percent of the

150 minute average 90-percent border. Since this is influenced the most by the second case, which is not performing well, this difference is not that high.

In general, we conclude that the intervals are performing well, especially compared with the current prognosis. This prognosis is based only on the failure form and would therefore give the same prognosis for case 1 and 2 and for cases 3, 4 and 5. These are significantly different as shown in Table 6.4.

Table 6.4: The Intervals determined by the Model compared with the actual, real, Situation and the Difference between the Two.

	50% Repaired	80% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Case 1 – Model	51	138	205	37%	51%	62%
Case 1 – Reality	49	118	223	32%	54%	71%
Case 1 - Difference	+2	+20	-18	+5%	-3%	-9%
Case 2 – Model	29	84	129	50%	66%	77%
Case 2 – Reality	23	45	81	53%	80%	84%
Case 2 - Difference	+6	+39	+48	-3%	-14%	-7%
Case 3 – Model	29	54	120	50%	77%	92%
Case 3 – Reality	22	49	88	60%	80%	88%
Case 3 - Difference	+7	+5	+32	-10%	-3%	+4%
Case 4 – Model	64	113	192	22%	42%	60%
Case 4 – Reality	60	116	212	9%	44%	56%
Case 4 - Difference	+4	-3	-20	+11%	-2%	+4%
Case 5 – Model	34	62	130	44%	71%	87%
Case 5 – Reality	28	67	133	40%	75%	80%
Case 5 - Difference	+6	-5	-3	+4%	-4%	+7%
Average Difference	5	14	24	6.6%	5.2%	6.2%

6.6. Conclusion

In this chapter we analyse the differences between the district control offices to a more detailed degree. This required values for variables not just from the database, but from the district control offices themselves and enables us to identify new variables influencing the FRT. Hereby, a more accurate prediction model is developed.

Visits to multiple district control offices result in four new variables influencing the FRT: movability around the path, age of the rails, a sharp corner and the PCA active in that area. The possible value these variables can take differs from a textual value ("Volker Rail", "Strukton") to a binary value (there is a sharp corner or not). These values influence the FRT in different ways and severities shown in section 6.3. Where the movability can either have a positive or a negative effect on the FRT, a really sharp corner is a sign that the FRT of a TOBS will be lower.

Applying the variables on the already known percentiles is done via four steps:

1. Determine the probability distribution and percentiles based on the failure form.
2. Apply infra-equipment type variables.
3. Apply locational variables.
4. Adjust the probability distribution and percentiles of step 1 based on the variables.

Based on these steps, with the example given in section 6.4 and the developed tool it is possible to determine a reliable initial prognosis for the failure recovery time of a certain infra-object.

Table 6.4 shows that this prediction by the model is fairly close to the actual situation: the success-percentage of the 25-minute blocks is on average only five percent off for the different cases.

7. Practical Application of the FRT Prognosis

This chapter explains the practical tests of the FRT prognosis method run in The Hague. It explains what has been done and the results of the first feedback circle. This will be the basis for multiple of these loops, making it possible to eventually develop a fully-functioning tool. During chapter 7, first the test run is explained, this is followed by the feedback and an overview of things still missing and how to acquire these in the future.

7.1. Execution of the Practical Test

Developing a perfect theoretical model is a good start, but it does not directly result in any performance improvements. To reach this, the theory should be applied in practise. How to apply this is researched in this chapter. This section explains how the first test was done.

As explained earlier in this research, district control office The Hague is used often as an experimentation area. For this reason, the first experiment was done in this area. The method of section 6.4 was used to develop a sheet with all the possible situations and their corresponding percentiles. This tool was explained and tested with the manager of the district control office The Hague. He, in turn, discussed it with several Trdl's (movement inspectors) and DVL's (decentralized traffic control). They discussed the application of the tool on hypothetical failures (the amount of real failures isn't that high to test it extensively based on real problems in a short amount of time). The results of these sessions are given in the next section.

7.2. Feedback on the Test

After discussing some theoretical failures, the district control office was capable of given some detailed feedback on the tool. The following most important positive and negative points emerged.

<u>Positive:</u>

- The format gave them quick results (in comparison with, say, a book);
- It improved the initial prognosis they could give about a failure;
- The decision whether or not to execute the first diagnosis block of 25 minutes is made easier for them.

<u>Negative</u>

- They have to manually select the values of the variables for a failure location and infra-equipment. This requires extra time in a situation requiring quick decisions.
- The probability of success of a failure recovery action is clear now, however, when to best execute a failure recovery action is not. This should also be offered.

Based on this first feedback cycle and the positive and negative points the ideal situation and the things missing in the current situation can be determined.

7.3. Comparison Ideal Situation vs. Current Situation

Based on the feedback from district control office The Hague, an accurate comparison between the ideal and the current situation of the practical application can be given.

In the ideal situation, the traffic control (VL) in the form of the Trdl's and DVL's have access to both the expected FRT and the amount of passenger hindrance when activating a WBI (the safety instructions per infra-equipment and thus, the tracks which go out of order when repairing a certain infra-equipment). They only have to fill in the time and the failed infra-object and failure form in order to obtain this. A representation of this is given in Figure 7.1. The example of a switch NIC is used to illustrate it.

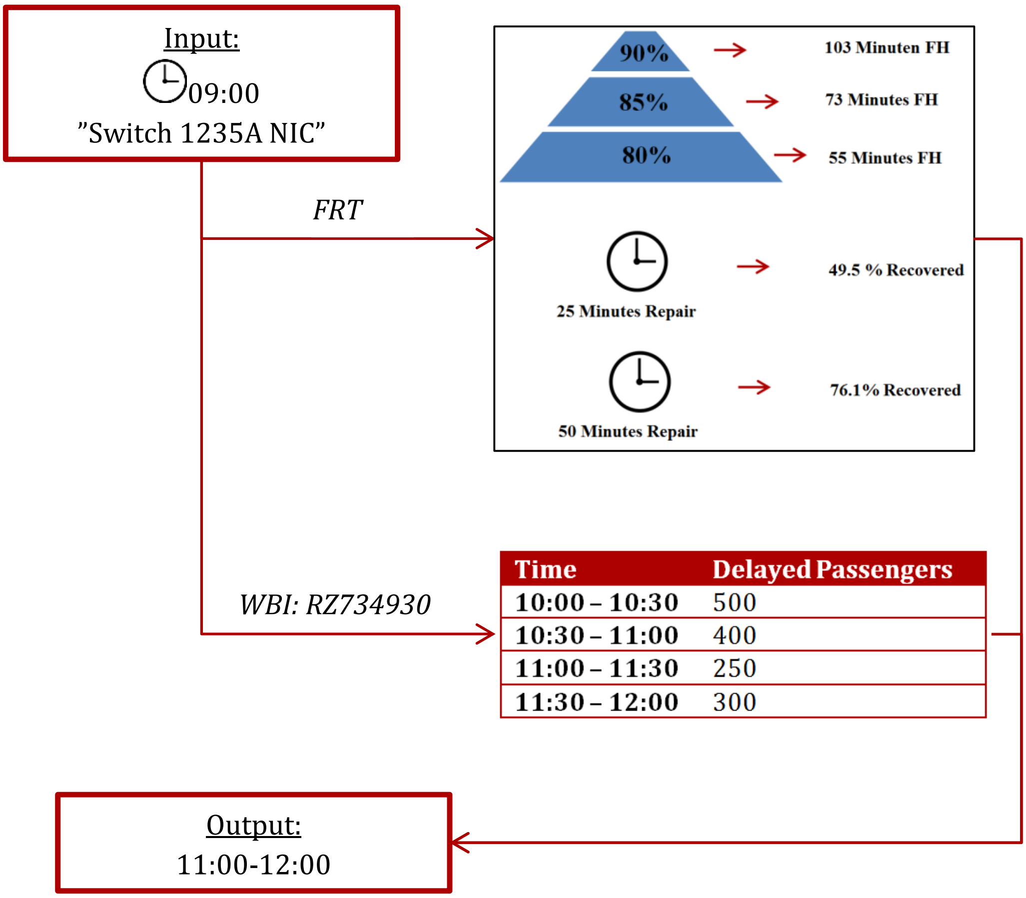


Figure 7.1: The Ideal Situation to Determine the Suitable Failure Recovery Moment for the Traffic Control.

This is the ideal situation, thus, a tool possibly developed over multiple years. Currently, only a few elements of this situation are present. This situation is shown in Figure 7.2.

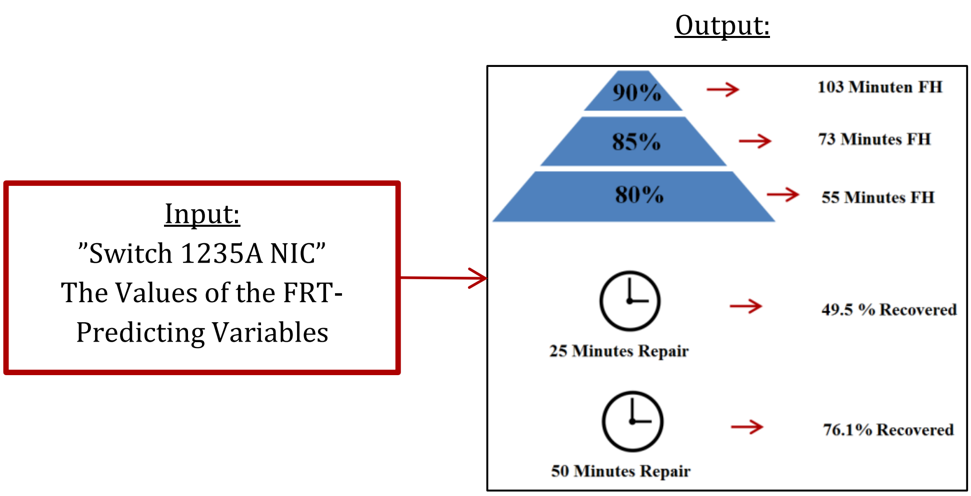


Figure 7.2: The Current Situation Offered by this Report.

When comparing Figure 7.1 and Figure 7.2, the elements which are not in the current situation can be discovered. The following elements are still missing:

- *Automatic generation of the values of the FRT predicting variables based on the infra-equipment type and the failure form.* In the ideal situation the VL only enters the infra-equipment and failure form and the tool automatically connects this with the correct values of the relevant variables. Thus, it knows, for instance, the PCA and the movability around the path. This saves time. In order to acquire this, the contractor responsible for the infra-equipment in the region has to be contacted. A session of one or two days with the infrastructure-experts of the contractor enables us to fill in every variable per infra-equipment. These data can be used to automatically fill in the missing values: a relatively easy task.
- *The delayed passengers per hour based on the WBI.* The second element missing is a bit more complicated. Determining the amount of delayed passengers based on the WBI of an infra-equipment depends upon multiple variables. This would therefore require a whole new research to be started. Fortunately, a first design for a tool was already developed by the company: CQM, hired by ProRail. Linking this research with that tool makes developing a first version of the ideal failure recovery moment-tool possible. The current version of this tool makes it possible to see the amount of delayed trains when a certain infra-equipment is no longer available. By manually selecting the infra-equipment's not available under the activation of a certain WBI, the WBI effect can be measured. Two things still missing, but being worked upon, are the effect of the time in which a WBI is activated and the actual amount of delayed passengers instead of the amount delayed trains. These are developed in the coming months.
- *The link between the two tools.* Linking these tools would result in the perfect interval to execute the failure recovery. Accomplishing this would require the help of an ICT-expert. This link and tool have to be constantly improved. This can be done by the Deming circle of Appendix J.

7.4. Conclusion

This chapter applies the model developed in chapter 6 to the practical situation in test-area The Hague. This allows the first feedback in the possible development of a failure recovery moment-tool to be analysed. During the first test, the percentiles for the failure recovery process are combined with the relevant variables and reviewed with the district control office.

The test reveals the need for the combination of the failure recovery time-model developed in this report with the expected passenger hindrance per hour. This enables a Trdl or DVL to select the best moment to execute a failure recovery based on its expected repair length and the amount of expected delayed passengers.

The two things still missing to develop this are the automatic variable value generation and the passenger hindrance per hour. The first can be determined with the help of the infrastructure experts and the contractor. CQM has already developed a first effort to construct a tool for the latter: the hindrance per hour.

The addition of these elements enables ProRail to develop a tool to determine the moment to execute a failure recovery with the least amount of hindrance for train passengers.

8. Conclusions and Recommendations

This chapter concludes the report. It starts with the main conclusions drawn from the research and the main limitations of this. Based on this, the end of the chapter shows the recommendations we make. Both the main research and the extensions towards the temporary failure recovery actions and the Deming circle are included.

8.1. Conclusions from the Research

Motivation: In this research we identify percentiles and influential variables considering the failure recovery time (FRT). A more accurate prediction of the FRT enables ProRail to supply better information to passengers (via the railway undertaking), to give the mechanic a more suitable interval to repair a failure and to support the traffic control in their decision process concerning the failure recovery moment (during the day or evening) and priority level.

The decision when to repair a failure is currently done by the traffic control based on experience. Traffic control makes a decision by taking three factors into consideration: the necessary WBI (the safety instruction) of the failure, the expected amount of passengers at this moment and in the close future and the expected failure recovery time. Making a careful consideration between these factors based on experience is prone to human errors. The wrong decision about the failure recovery moment is therefore often made. This research improves and simplifies this decision making by clarifying one of these factors: the expected failure recovery time. This is the first step made towards an automated and mathematically enhanced decision making process concerning the failure recovery time.

Current Prognosis: The current initial determination of the FRT is done based only on the failure form and comes in the form of one value: the average. This value is inaccurate in a lot of situations: for a TOBS, the prognosis is, on average, 24% too low while the prognosis is 25% too high for a failure concerning the status of the path. The performance of the prognosis is also not consistent: 72% of the FRTs of a level crossing failure are lower than the prognosis while this value is 62% for a TOBS-failure. We therefore conclude that the performance of the current prognosis has to be improved and made more consistent. Moreover, the amount of data far lower or higher than the prognosis showed that using a one-point prognosis is not a suitable method. We therefore used an interval-prognosis in the remainder of the report.

Prediction Methods: Several methods to predict the value of a variable are offered by the literature ranging from simple, inaccurate methods to extensive, difficult accurate methods. We evaluated several methods on accuracy, speed, operation simplicity, development flexibility and development practicality. Some methods turned out not be accurate enough (smoothing methods) while others aren't practical in this situation (neural networks). Four methods initially show promise to predict the FRT: confidence intervals based on probability distribution, regression analysis, prediction by the expert and nearest neighbour. These methods are applied on the data set.

The expert is capable of correctly predicting the order of longest to shortest FRT for the failure forms. However, he is not capable of giving a quantitative prediction of the FRT per failure form. This method is therefore not suitable for the FRT prediction model. Nearest neighbour is also not a suitable method. The difference between the predicted FRT by this method and the actual FRT ranges from an underestimation of 25 minutes to an overestimation of 75 minutes for specific cases. The most suitable method to base the initial prediction of the FRT on proved to be a combination of confidence intervals and regression analysis.

Based on the probability distributions of the failure forms, confidence intervals in the form of percentiles and percentage of failures recovered within a certain amount of 25-minute blocks are determined. This shows large differences between the failure forms and their repair duration. Where only about 40 percent of the path TOBS are repaired within 25 minutes, this is 80 percent for a sign turned off. The probability distribution also show the increasing repair probability of switch failures: the second 25-minute block is more effective than the first and the third is, in turn, more effective than the second.

The performance of the intervals can be further improved by applying relevant variables determined based on regression analysis. This analysis identified a large difference between the district control offices: a failure in the region of Groningen or The Hague is solved 30 minutes faster than a failure occurring in the regions: Kijfhoek, Roosendaal or Utrecht. Since no direct explanation for this difference can be given by ProRails employees or the database, we did some field research. With this practical research, new, more specific, variables are determined and quantified.

These, and their quantified effects are:

- *PGO contract*: A PGO contract lowers the expected FRT of a switch failure with eight minutes in comparison to the OPC-contract.
- *PCA active in the region.* The PCA has a significant influence on the FRT by its knowledge and its spare part strategy. The quantified effects of the PCA's on the FRT of certain failure forms are given in Table 6.1.
- *Grinding train*: The recent passing of a grinding train lowers the expected FRT of a path TOBS with thirty minutes.
- *Movability around the path*: A bad or good movability around the path respectively increases or decreases the FRT of a path TOBS with twelve or ten minutes.
- *Age of the path:* An old path decreases the expected FRT of a path TOBS with fourteen minutes.
- *Sharp corner*: A sharp corner just before the path with a TOBS decreases the FRT with thirty minutes.

The combination of the probability distribution and the effect of the relevant variables result in the expected FRT. Combining the two is done by adjusting the scale-value of the distribution based on the initial value and the value of the relevant variables. This new distribution is used to generate the intervals for this specific situation.

The evaluation of the model shows a good performance of it, especially when compared with the original prognosis. For instance: the success-percentage of the 25-minute blocks is on average only five percent off compared with the actual situation.

<u>Practical Application:</u> The practical tests of the FRT prediction model in The Hague show the willingness of the traffic control to use the model. The positive feedback also showed the necessity of such a tool. Traffic control moreover indicates a possible extension of the tool: the combination of the FRT prediction tool with the amount of passengers hindered per hour. This combination enables the traffic control to choose the best start time and the correct end time to

enable a mechanic to repair the failure with the least amount of passenger hindrance. A traffic control employee thus knows exactly when to postpone a failure recovery (and to which time) and when to repair a failure as quickly as possible.

Not every aspect of this ideal failure recovery moment tool is yet present. Now, every relevant regression variable has to be entered manually. It is possible to automate this by entering the value of every variable in the database and linking it to each infa-object. Another aspect not yet present is the amount of passenger hindrance at a certain time when activating a certain WBI (the safety instructions). This aspect is currently being researched by a department at ProRail and will therefore be present in the close future. The last missing aspect is the combination between the two tools. Multiple options for this are possible, however, we recommend to use an ICT-tool to combine the two.

<u>Temporary Failure Recovery:</u> Not every failure recovery action results in a fully functioning infra-object. It is also possible that a repair action is executed but it is only a temporary solution (see Appendix I). Thus, the object functions again but a full repair action is still necessary to make sure it continues functioning. About 17 percent of all the failure recovery actions executed are temporary failure recovery actions. These are, in general, executed for two reasons: first, because a temporary recovery is quicker than a full recovery and second, because there is no spare part present.

This research shows that temporary failure recovery takes, on average half an hour less than a full recovery (the recovery actions requiring the placement of a "slipkabel" and the replacement of a level crossing barrier are excluded). Thus, when the failure is severe but the recovery action is impractical, a temporary repair action now and a full recovery action during a more suitable time is an attractive action to spare half an hour now. When the spare part is missing the delivery time of the spare part is, on average, 150 minutes. Executing a temporary recovery action is therefore usually a good solution instead of waiting for the spare part for a full recovery.

The possibility to execute a temporary failure recovery action and its FRT is one of the options for the traffic control to choose from in the failure recovery moment tool.

<u>Deming Circle of Continuous Improvement:</u> The Deming Circle (the plan-do-check-act circle) applied in Appendix J has to be used to evaluate the decision made considering the failure recovery moment by traffic control. The wrong decision being made results in unnecessary delay of trains. Traffic control has to make a careful consideration between the FRT, the time till the evening, the influence of the WBI and the current influence of the failure on the passengers. Based on the actual and the expected values of these variables the choice made by the traffic control, the argumentation of this choice and the performance of this can be analysed afterwards.

A frequent execution of this circle will, in combination with the tool stated in the last chapter, continuously improve the decision making process of traffic control.

8.2. Limitations of the Research

This research is conducted while following the basic scientific research-'codes' considering qualitative and quantitative research. However, for several reasons (for instance: lack of time or practicality-issues) the research contains some small limitations stated in this paragraph. This enables future research to take the reliability of this research into account and, possibly, enhance this.

- An important limitation of the research is the use of data from only three years due to the lack of reliable data from earlier on. For some of the failure forms the amount of data is therefore relatively low for a reliable conclusion. This limitation thus mainly applies on the failure forms with a low occurrence. It should be kept in mind that the results of the research do not include the yearly changes in failure recovery time (FRT) performance. Thus, the results can't be applied on data far later than 2016 without carefully checking it.
- Some of the variables introduced in the latter chapters are up to human interpretation. We did attempt to narrow the effect of these interpretations down by carefully stating the possible status of the variables and their implications. However, some personal interpretation still persists.
- The prediction of the FRT does not contain the (initial) travel time of the mechanic to the failure location. Since we initially focused upon the required repair interval for a mechanic during the night, instead of focusing on the extension towards a priority-determination tool, this is not taken into account. For the purpose of this research, this is not a problem since the role of the traveling time is different (a restriction of the possible start of the repair instead of the necessary time that a WBI, the safety instructions, are activated). However, this limits the use of the research when traffic control wants to use it for passenger information about the duration of a failure from that moment on.

8.3. Recommendations

After six months of theoretical and practical research and the application of the results we make the final recommendations. These comprise suggestions directly related to the research but also suggestions about some of the general processes at ProRail (like the collection of data). The main recommendations we make are:

- The current prognosis (an average per failure form taken years ago) should be replaced by the FRT-prognosis method from this report. As shown in chapter six, this prognosis outperforms the current prognosis both on accuracy and practicality. Thus, the prognosis has to be based on the relevant variables and probability distributions from this research.
- The determination of the priority level and possibly: the repair moment, the length of the recovery interval for a mechanic and the information given to the passenger concerning the duration of failure has to be based on this prognosis. This enhances the accuracy and results of a decision.
- We recommend to discuss the results of this research with the experts of the department: architecture and technique to further enhance the prediction by acquiring new variables. By showing the results of the research this department, and their experts, they are more willing to further research variables influencing the FRT. This enables ProRail to enhance the accuracy of the prediction even further. This also serves to show the department how to lower the FRT by selecting infra-layouts with a favourable value on certain variables.
- The current data collection and analysation is difficult, prone to errors and inefficient due to the high amount of textual data. Improving this by adding more numerical and standard input-values is highly recommended for further successful datamining research.

- We recommend to update the values and the impact of the values yearly. Due to changes in the layout of the infrastructure, an improvement or deterioration of the performance of the contractor or other changes affecting the duration of the FRT the predictions accuracy will go down. Keeping it updated prevents this. Furthermore, when a large change takes place, for instance: a change of contractor in a certain area or a new station opening, the factors in this area have to be researched again.
- We further recommend to optimize the FRT prediction tool by automating the variable-input process. By consulting the experts of the contractor, the values of the relevant variables per infra-equipment can be determined beforehand, saving the traffic control time when a failure occurs. Furthermore, the combination with the CQM-tool, predicting the amount of passenger hindrance per failure, is essential to be able to select the repair moment and interval with the least amount of delays for the passengers: an important goal of ProRail.
- It is important to let the TRDL and the DVL reflect upon their decisions concerning the failure recovery moment. This can best be done by making use of the Deming circle (see Appendix J) which aims to continuously improve a process.
- Include the temporary FRT (see Appendix I) in the tool selecting the right failure recovery moment (FRM). This adds a further options to the tool which represents the option with the least amount of passenger hindrance in certain situations.
- We recommend to use the following roadmap to implement the FRM-tool. This gives a stepwise method to implement the tool at every district control office and identifies the responsible parties for every step.

Roadmap- Failure Recovery Moment - Tool

1. Automatically fill in Values for the FRT-Variables.	**Who?** Contractor and district control office The Hague. **How?** An infrastructure-expert of the contractor and the district control office will go through the infra-objects one by one. In the span of one or two days they can determine the values of the infra-oject for every relevant variable.Implement this.
2. Receive and Use fFirst Feedback.	**Who?** Traffic control of district control office The Hague and coordinator tool. **How?** With a feedback session the traffic control can inform the coordinator of the tool about the limitations of the tool. These are analysed and improved.
3. Extend to Multiple District Control Offices.	**Who?** District control offices, contractors and coordinator tool. **How?** At every district control office, the need and relevancy of the tool will be clarrified. After this, the values for the variables of the infra-objects have to be clarified again. Next, the tool can be introduced one office at a time.
4. Develop Passenger Hindrance Tool.	**Who?** CQM and project leader hindrence-tool **How?** The current influence tool has to be extended with the WBI-influence and the hourly amount of delayed trains/passengers.
5. Combine Passenger Hindrance with FRT-Prediction.	**Who?** CQM, coordinator tool and project leader hindrence-tool. **How?** The objectives of the tool have to be clarified first. CQM or ProRail experts can next combine the seperate tools to fulfill the objectives.
6. Implement Combination at one District Control Office.	**Who?** District control office The Hague and coordinator tool. **How?** With the help of clear courses and instructions the use and value of the tool can be introduced at the office. Afterwards, an expert has to be present in order to help traffic control with the tool and to anser questions.
7. Receive and Use First Feedback.	**Who?** District control office The Hague, CQM and coordinator tool. **How?** The first feedback is received from the control office. CQM and the cooridinator of the tool together verify, analyse and improve the tool based on this.
8. Extend the tool to Multiple District Control Offices	**Who?** District control offices and coordinator tool. **How?** The need and relevany of the tool have to be explained at every district control office. Again, clear courses and instructions are used to introduce the tool one office at a time to the rest of The netherlands. Attention should be paid to feedback and questions about the tool.

Bibliography

Berson, A., Smith, S., & Thearling, K. (2000). An Overview of Data Mining Techniques. In A. Berson, S. Smith, & K. Thearling, *Building Data Mining Applications for CRM.* New York: McGraw-Hill Companies.

Brook, Q. (2014). *Lean Six Sigma and Minitab (4th Edition).* Winchester: OPEX Resources Ltd.

BusinessDictionary. (2015). Retrieved September 10, 2015, from Mean Time to Repair Definition.

Cadwallader, L. (2012). *Review of Maintenance and Repair Times for Components in Technological Facilities.* Idaho Falls: Idaho National Laboratory.

Colquitt, J., & Zapata-Phelan, C. (2007). Trends in Theory Building and Theory Testing: A Five-Decade Study of the Academy of Management Journal. *Academy of Management Journal, 50 (6)*, 1281-1303.

Cordeau, J. -F., Gendreau, G., Laporte, G., Potvin, J.-Y., & Semet, F. (2002). A Guide to Vehicle Routing Heuristics. *Journal of the Operational Research Society 53*, 512-522.

Craparo, R. (2007). Signifiance Level. In N. Salkind, *Encyclopedia of Measurement and Statistics 3* (pp. 889-891). Thousand Oacks: CA: SAGE Publications.

Dasarathy, B. (1991). *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques.* IEEE Computer Society.

Dhillon, B. (2002). *Engineering Maintenance: A Modern Approach.* Boca Raton: CRC Press LCC.

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226-231.

Fox, A. (2002). *Towards Recovery-Oriented Computing.* Stanford University.

Glover, D., Jenkins, W., & Doney, S. (2008). *Least Squares and Regression Techniques Goodness of Fit and Tests. Non Linear Least Squares Techniques.* Woods Hole: Woods Hole Oceanographic Institute.

Goodell, R. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series.* Englewood Cliffs: Prentice-Hall.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques.* Waltham: Morgan Kaufmann.

Heerkens, H., & van Winden, A. (2012). *Geen Probleem.* Nieuwegein: Van Winden Communicatie.

Ibrahim, J., Chen, M., Lipsitz, S., & Herring, A. (2005). Missing-Data Methods for Generalized Linear Models. *Journal of the American Statistical Association, Vol.100, Is. 469*, 332-346.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods and Algorithms.* New Jersey: John Wiley & Sons.

Klutke, G., Kiessler, P., & Wortman, M. (2015). A Critical Look at the Bathtub Curve. *IEEE Transactions on Reliability 52(1)*, 125-129.

Lewis, E., & Schwartz, S. (2004). *Sea Salt Aerosal Production: Mechanisms, Methods, Measurements and Models.* Washighton DC: American Geophysical Union.

Markowski, C., & Markowski, E. (1990). Conditions for the Effectiveness of a Preliminary Test of Variance. *The American Statistician 44 (4)*, 322-326.

Mendehall, W., Sincich, T., & Boudreau, N. (1995). *A Second Course in Statistics: Regression Analysis (Vol 5).* New Jersey: Prentice Hall.

NASA. (2002). Mean Time to Repar Predictions. *NASA Maintenance Technique AT-2*, pp. 7-12. Retrieved from NASA Technical Memorandum.

Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society Vol. 236, Issue 767*, 333-380.

Oosterbaan, R. (1994). Frequency and Regression Analysis of Hydrologic Data. In H. Ritzema, *Drainage Principles and Applications, Publication 16* (pp. 175-224). Wageningen: International Institute for Land Reclamation and Improvement (ILRI).

Rahm, E. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Data Eng. Bull. 23.4*, 3-13.

Ross, S. (2009). *Introduction to Probability and Statistics for Engineers and Scientists (4th ed.).* Gent: Academic Press.

Rother, M. (2010). *Toyota Kata.* New York: MGraw-Hill.

Sarkar, J., & Chaudhuri, G. (1998). Availability of a System with Gamma Life and Exponential Repair Time under a Perfect Repair Policy. *Statistics & Probability Letters 43*, 189-196.

Song, Y., Tobagus, W., Raymakers, J., & Fox, A. (2004). *Is MTTR More Important Tahn MTTF For Improving User-Perceived Availabiility.* Stanford University.

Walck, C. (2007). *Hand-book on Statistical Distributions for Experimentalists.* Stockholm: University of Stockholm.

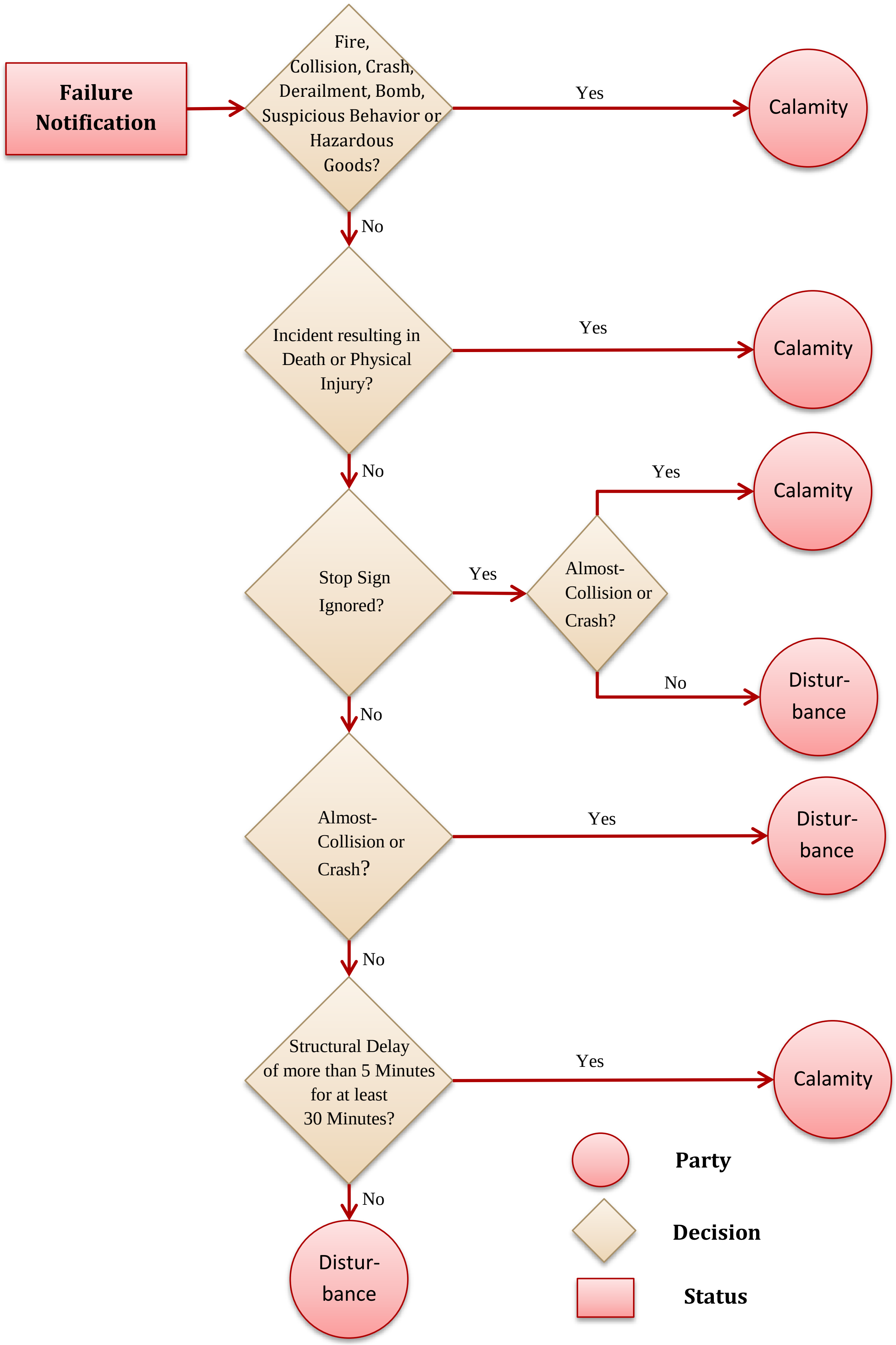
Wohl, J. (1982). Maintainability Prediction Revisited: Diagnostic Behaviour, System Complexity and Repair Time. *IEEE Transactions on Systems, Man and Cybernetics Vol. 12 No. 3*, 241-250.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data Preparation for Data Mining. *Applied Artificial Intelligence: An International Journal. Vol. 17, Issue 5-6*, 375-381.

Appendix A: Detailed Research Planning

Research Question*	Expected Time	Start	End
1.1	5 Days	21-09	25-09
1.2	2 Day	28-09	29-09
1.3	11Days	30-09	14-10
1.4	2 Days	15-10	16-10
2.1	12 Days	19-10	03-11
2.2	2 Days	04-11	05-11
3.1	10 Days	06-11	20-11
3.2	4 Days	23-11	26-11
3.3	4 Days	27-11	02-12
3.4	3 Days	Depends on date meeting contractor	
4.1	2 Days**	03-12	04-12
4.2	4 Days***	07-12	10-12
4.3	3 Days	11-12	15-12
5.1	7 Days	16-12	24-12
5.2	3 Days	04-01	06-01
5.3	4 Days	07-01	12-01
5.4	12 Days	13-01	28-01
6.1	15 Days****	29-01	19-02
6.2	8 Days	22-02	02-03
***This planning was made during the production of the plan of approach. Since the research questions have been updated later on, this planning can't be read directly via the research questions, only by start- and end-date.**			
****Can partially be done during 1.1**			
*****Can partially be done during 1.3**			
******Length of test depends on research-time remaining**			

Appendix B: Decision Tree Failure Type



Appendix C: TIS's and Corresponding FRT prognosis

TIS-type	Description	Prognosis
5	Notification / discovery bomb	100
4	Hazardous substances	120
3	Collision cyclist	145
3	Collision big animal	80
3	Collision object	120
3	Collison person (injured)	90
3	Collision person(dead)	145
3	Collision shunt object	360
3	Collision train-train	600
3	Collision buffer stop	240
3	Collision viaduct/ bridge	80
2	Fire/smoke bank	50
2	Fire/smoke building/tunnel	60
2	Fire/smoke material	100
1	Calamity abroad	120
1	Extreme weather	220
1	Hindrance because of government help services	80
1	Hindrance logistical problem	70
1	Hindrance object, vehicle/animal	110
1	Hindrance persons railroad	40
1	Hindrance passengers/personnel	40
1	Hindrance vandalism	180
1	Infra - steering/communication	130
1	Infra- TOBS	130
1	Infra – catenary	360
1	Infra – bridge defect/disturbed	100
1	Infra – defect	90
1	Infra – level crossing / sign-failure	120
1	Infra – section-failure	120
1	Infra – power outage	120
1	Infra – status railroad	210
1	Infra – switch	120
1	Unplanned repaired activities	180
1	Stranded train (defect)	70
1	Stranded train (other)	70
1	Repair activities longer than expected	110
1	Remaining	60

Appendix D: Explanation of Failure Forms

Failure Form	ProRail Notification	Explanation
Path TOBS	TOBS in sectie spoor	There is a TOBS (train vacancy detection failure) in a section with only paths (thus, not a section with a switch/level crossing).
Switch NIC	Wissel NIC	A switch is NIC (can't be controlled).
Level crossing disturbed	Overweg gestoord	Level crossing is disturbed. When a level crossing barrier remains closed for five minutes this failure notification is automatically generated.
Level crossing collision	Aanrijding overweg	Someone drove into the level crossing. This also includes the demolition of, for instance, a barrier on purpose since the results are the same.
Switch TOBS	TOBS in wissel	A TOBS in a section with a switch in it (the TOBS is usually caused by the switch and thus it is easier to find the cause).
Sign Out of Order	Sein irregulier	A sign showing that a machinist has to stop is behaving irregular. For instance: it turns on and off or does not respond to the controls.
Level crossing barrier not responding properly	Overwegboom functioneert niet volledig	The barrier of the level crossing is not functioning properly. For example: it closes only half or opens too late.
Other failures	Overige storingen	The failures which do not fit into one of the other categories.
Path subsidence	Spoorverzakking	The path has subsided usually resulting in a bump felt by the machinist.
Switch disturbed	Wisselstoring	The most general term when a switch is disturbed.
Status path, others	Toestand spoor, overig	The status of the path is not perfect but this is not a buckling or a subsidence. For instance: a fracture of the path.
Level crossing specific damage	Overwegstoring, overig	A failure of the level crossing which is not a TOBS or a general disturbance but more information about the failure in known.
Buckling in path	Knik in spoor	When two path pieces are not perfectly lined up.
Sign turned off	Sein gedoofd	The light of the sign suddenly turns off (and can't be turned on again).
Smouldering sleepers	Smeulende bielzen	The sleepers are smouldering
Level crossing TOBS	TOBS in overweg	A TOBS in a section with a level crossing in it (the TOBS is usually caused by the level crossing and thus it is easier to find the cause).
Bridge disturbance	Brug gestoord	A bridge (with paths on it) is disturbed.
Overpass collision	Aanrijding viaduct	Someone (usually a lorry) drove into an overpass with a path on it.

Failure Form	ProRail Notification	Explanation
Switch, specific damage	Wissel, specifieke schade	A failure in a switch where the specific damage/cause is known in advance.
ATB-system failure	ATB-systeem storing	The ATB-system ("Automatische Trein Beïnvloeding"), the system which can influence the speed of a train to prevent it from speeding, is disturbed.
Fire	Brand	A fire
Cable damage	Kabel schade	Cables which are not from one of the infra-equipment mentioned above are damaged/stolen.

Appendix E: Pearson's Chi-Squared Test

Pearson's chi-squared test is specific part of the goodness-of-fit-tests. Goodness-of-fit-tests describe how well a statistical model describes a certain set of data points. This can, for instance, be used to determine whether two samples are drawn from the same distribution. Pearson's chi-squared test is subgroup of these tests. This test is a statistical method testing how well an assumed probability distribution fits a set of data points. When a set of data points and a distribution pass this test, it is safe to assume that the distribution fits the points. During the test, the assumed distribution is used to determine the amount of data points expected within a certain interval. This is compared with the actual amount of data points within this interval. Based on formula E.1, O_i, the actual amount of data points in interval i, and E_i, the expected amount of data points in interval i, the test statistic χ^2 is calculated. When the test-statistic is lower than the corresponding value of the chi-square distribution there is no reason to reject the assumed probability distribution (Glover, Jenkins, & Doney, 2008).

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \text{(E.1)}$$

The practical application of Pearson's chi-squared test is illustrated in the following example. During this example the assumed probability distribution for the failure form: level crossing collision is tested. The 3-parameter Weibull-distribution was suggested by Minitab (the program used to suggest the probability distribution). The 3-parameter Weibull-distribution is a continuous probability distribution related to multiple other distributions and is often used to model the time until a technical device fails. In this case the assumed shape parameter β, the parameter affecting the shape of the distribution, is 1.14. The scale parameter α, the parameter determining how spread out the distribution is, is 77.30. The threshold t_0 , the shift of the distribution, is -0.13. Whether this probability distribution fits the data points well is tested next in the next six steps.

Step 1: Determine Amount of Intervals. The amount of intervals in which the expected and observed data points are distributed. A common rule of thumb for the amount of intervals is to take the square root of the amount of data points. This is also used in this case. There are 165 data points. The square root of this is 12.85. Rounding this results in a used of thirteen intervals.

Step 2: Determine Expected Amount of Data Points per Interval. Since there are thirteen intervals and 165 data points, the expected amount of data points per interval is 165/13 = 12.69.

Step 3: Determine Borders of Interval. In this step the border per interval is determined. These borders can be used to determine the amount of data points which fall into each interval. The border is determined based on the assumed probability distribution and the amount of intervals. For example: the first interval. Since there are thirteen intervals, 1/13 = 7.69% of the data points should fall into this interval. The inverse Weibull (formula E.2) combined with the parameters and the percentage of data points that should be under this interval-border (x) results in the border. Thus, for the first interval x is 0,0769 while x is 4 times 0,0769 is 0,308 for the fourth interval.

$$Q(\alpha, \beta, x) = \alpha * (-\ln(1-x))^{1/\beta} \qquad \text{(E.2)}$$

This results in the border per interval as shown in Table E.1.

Table E.1: The Borders corresponding with the Intervals used in Pearson's Chi-Square Test for the Failure Form: Level Crossing Collision.

Interval	Border
1	8,59
2	16,24
3	24,06
4	32,30
5	41,17
6	50,91
7	61,83
8	74,41
9	89,41
10	108,24
11	134,03
12	176,61
13	∞

Step 4: Determine Test Statistic Value. Formula E.1 can be used to determine the test statistic. First, the amount of data points which fall in each interval based on the borders per interval (Table E.1) are determined. This amount is compared with the expected amount of data points per interval (see step 2) via formula E.1. Via this formula, the value of the test statistic χ^2 is 7.62.

Step 5: Determine Chi-Square Value. The chi-square value depends upon two factors: the degrees of freedom and the statistical significance value. The degree of freedom is calculated by subtracting the amount of intervals (thirteen) by the amount of parameters estimated (three: the scale-, shape- and threshold-parameter). Thus, there are ten degrees of freedom. For the statistical significance- level a value of 0.05 is used. These two values results in a value of 18.31 for the chi-square distribution.

Step 6: Compare Test Statistic with the Chi-Square Value. The test statistic is lower than the chi-square value (7.62 compared to 18.31). It is therefore safe to assume that the assumed probability distribution fits the data points. Thus, the 3-parameter Weibull distribution with the parameters: 1.14;77.30 and -0.13 can be used for the failure form: level crossing collision.

Appendix F: Detailed Confidence Intervals for the Failure Forms: TOBS Path and Level Crossing Disturbed

Just like the failure form: switch NIC, it is possible to extract certain smaller failure forms out of the failure forms: path TOBS and level crossing disturbed, however, to a smaller extent. In this appendix, the same steps and tables are introduced as in the confidence interval section in chapter 4.2.2. First, the failure form: path TOBS is treated, after this the failure form: level crossing disturbed is treated.

During the failure form: path TOBS a TOBS in a specific section of a path occurs. There are different types of sections like the already introduced switch and level crossing sections. In this appendix, two more sections are identified. A TOBS in one of these sections does not occur often, however, often enough to base a reliable prediction on. Besides this, they behave relatively different than the other path TOBS' and thus influence the confidence intervals. The two new sections are the following:

- Assenteller Section TOBS. An assenteller section is a section containing an assenteller. An assenteller counts the amount of passing axles by measuring the disturbance in its magnetic field by a passing wheel. There are always two assentellers in a certain section. When both of them counted the same amount of axles, the system knows the train has passed this section. A large advantage of this method is possibility to remove the, susceptible to failure, already introduced ES-joints.
- Psssl Section TOBS. Psssl stands for "prikspanning-spoorstroomloop" (lightly translated as: peak-voltage electrical current. This device helps detecting a train by occasionally sending out a pulsing higher voltage (instead of the normal, non-pulsing, relatively low voltage). By doing this, rust on the tracks does no longer influence the current to the same extent as it did before.

Again, Minitab is used to determine the most suitable probability distribution. This is done not only for the two mentioned sections, but also for the remaining regular path TOBS. The suggested distributions are shown in Table G.1.

Table G.1. The Assumed Probability Distributions and their Parameters of the New TOBS Failure Forms.

Failure Form	Assumed Distribution	Shape	Scale	Location	Threshold
Regular Path TOBS	3-Parameter Weibull	0.86	113.16	-	-0.10
Assenteller Section TOBS	Exponential 1 Parameter	-	51.58	-	-
Psssl Section TOBS	Weibull	0.80	54.77	-	-

The goodness of fitness is applied on the assumed distribution. As can be seen from Table G.2., all of the test statistic values are well below the chi-square boundary value. Thus, it is safe to assume that the suggested distributions are a good representation of the data points.

Table G.2: The Result of Pearson's Chi-Square Test for all the Failure Forms and its Suggested Distribution.

Failure Form	Assumed Distribution	Clusters	Degrees of freedom	Test statistic	Chi-square value
Regular Path TOBS	3-Parameter Weibull	22	19	17.52	30.14
Assenteller Section TOBS	Exponential 1 Parameter	6	5	4.03	11.07
Psssl Section TOBS	Weibull	6	4	8.03	9.49

For the failure form: level crossing disturbed, it is possible to subtract the new failure form: detection system level crossing disturbed out of it. This failure form includes a disturbance in the detection system which signals to a level crossing that a train is coming or has passed. A failure in this system will therefore result in a level crossing not closing or not opening since it does not get the signal to do this. Besides this failure form, the level crossing failure form also persists, but now without the detection system. The assumed probability distribution is shown in Table G.3.

Table G.3: The Assumed Probability Distributions and their Parameters of the New Level Crossing Failure Forms.

Failure Form	Assumed Distribution	Shape	Scale	Location	Threshold
Level crossing disturbed	Weibull	1.11	45.49	-	-
Detection Level Crossing Disturbed	3-Parameter Weibull	0.89	92.52	-	-1.39

Table G.4 shows the goodness of fit-test applied on the assumed probability distribution and the data points. Again, there is no reason not to accept the suggested distribution.

Table G.4: The Result of Pearson's Chi-Square Test for all the Failure Forms and its Suggested Distribution.

Failure Form	Assumed Distribution	Clusters	Degrees of freedom	Test statistic	Chi-square value
Level crossing disturbed	Weibull	14	11	17.47	19.68
Detection Level Crossing Disturbed	3-Parameter Weibull	7	4	5.32	9.49

Based on the assumed probability distributions of the failure forms, it is possible to develop new intervals. These are shown in Table G.5.

Table G.5. The Intervals resulting from the New Failure Forms for a Path TOBS and a Level Crossing Disturbed.

Failure Form	50% Repaired	80% Repaired	85% Repaired	90% Repaired	< 25 Minutes	<50 Minutes	< 75 Minutes
Regular Path TOBS	57.9	177.6	219.2	279.5	36.1%	48.2%	57.5%
Assenteller Section TOBS	19.8	63.5	78.4	99.3	57.8%	74.0%	83.0%
Psssl Section TOBS	18.6	80.1	103.0	136.8	57.2%	70.1%	78.6%
Level crossing disturbed	16.7	50.4	61.6	77.0	62.3%	79.8%	89.5%
Detection Level Crossing Disturbed	46.7	139.8	171.9	218.1	39.8%	53.3%	63.4%

As can be seen from the new intervals, introducing the new failure forms has a significant input on the old failure forms. Since some easier to repair failure forms are now removed from the path TOBS failure form, the intervals to repair a certain percentage of the failures get even larger. This is the other way around for the level crossing disturbance. Since the detection system takes longer to repair, removing it from the failure form makes repairing it even more likely after a certain amount of time.

Appendix G: Confidence Intervals per Region and Failure Form

In this appendix the 80%, 85% and 90%-confidence intervals and the percentage of failures recovered after 25, 50 and 75 minutes are given based on the district control office a failures occurs in (the 13 separate district control offices in The Netherlands each take care of the care in a specific area of the country). This is determined by the regression analysis of chapter 4.2.2. Per failure form (starting with the highest occurring failure form) the intervals and percentages are given per district control office. The values are determined by combining the confidence intervals based on the probability distribution (Table 4.3 and 4.5) with the regression formulas of Table 4.8. The interval values of Table 4.5 are the initial values of the intervals. These are increased (or decreased) by the performance of the district control office in comparison with other district control offices on the FRT of that failure form. This performance of a district control office is determined by comparing it with the weighted average performance. This average performance can be determined by adding the multiplication of the occurrence of the failure form per district control office and the performance factor of Table 4.8.

For instance: 9.15% of the path TOBS are from district control office Amsterdam, 10.33% from district control office Eindhoven and so on. The absolute performance for each of these district control offices are shown in the equation of table 4.8. For district control office Amsterdam, this is -24.0 and for district control office Eindhoven this is -7.8. By adding the multiplication of these two factor of all four regions the average performance is -10.6. Since Amsterdam has a factor of -24.0, it has a 13.4 minute better performance on this failure form compared to other regions. Meanwhile, Eindhoven has a 2.8 minute disadvantage. These numbers are added to the intervals of Table 4.5 to show the effect of a failure occurring in a specific region. Thus, usually 80 percent of the path TOBS are repaired after 122.6 minutes, but in the region of district control office Amsterdam 80 percent is repaired within 122.6-13.4 = 109.2 minutes.

With the same reasoning the percentage of data points repaired within a certain time can be determined. Now, the probability distribution of Table 4.3 is used combined with the just determined difference per district control office per failure form. The assumption is made here that the probability distribution of a specific failure form does not differ between different district control offices.. This is a safe assumption since the process of failure recovery stays the same with the same failure form, the difference between regions is mainly a constant factor (for example: the distance to the voltage cabinet at the station).

Path TOBS

Path TOBS	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	133.2	89.5	158.2	109.2	125.4	61.3	92.2
85%	158.6	114.9	183.6	134.6	150.8	86.7	117.6
90%	194.4	150.7	219.4	170.4	186.6	122.5	153.4
<25 Min	31.8%	58.4%	9.5%	48.1%	37.6%	69.8%	57.2%
<50 Min	48.6%	68.7%	31.8%	60.9%	53.0%	77.2%	67.7%
<75 Min	61.3%	76.4%	48.6%	70.5%	64.6%	82.9%	75.7%

Path TOBS	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	160.2	100.3	163.2	161.2	156.4	101.0
85%	185.6	125.7	188.6	186.6	181.8	126.4
90%	221.4	161.5	224.4	222.4	217.6	162.2
<25 Min	7.5%	53.0%	4.3%	6.4%	11.4%	52.7%
<50 Min	30.3%	64.6%	27.9%	29.5%	33.2%	64.3%
<75 Min	47.5%	73.3%	45.7%	46.9%	49.7%	73.1%

Extra factors: Subtract 39.5 Minutes of the FRT if the grinding train has passed recently.

Switch NIC

Path TOBS	
80%	74.3
85%	83.9
90%	96.5
<25 Min	36.5%
<50 Min	62.1%
<75 Min	80.4%

Extra factors: Subtract 20.7 minutes of the FRT if the contractor has a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 10.3 minutes can be subtracted from the above table for an PGO-contract.

Level Crossing Disturbed

Level Crossing Disturbed	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	64.7	57,1	79.7	59.4	58.9	37.8	20.2
85%	78.1	70,5	93.1	72.8	72.3	51.2	33.6
90%	96.9	89,3	111.9	91.6	91.1	70.0	52.4
<25 Min	53.5%	60,4%	36.5%	58.4%	58.8%	73.7%	81.9%
<50 Min	72.6%	76,7%	62.3%	75.5%	75.8%	84.6%	89.5%
<75 Min	83.9%	86,4%	77.8%	85.7%	85.8%	91.0%	93.9%

Level Crossing Disturbed	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	35.1	31.4	48.0	35.0	43.4	81.5
85%	48.5	44.8	61.4	48.4	56.8	94.9
90%	67.3	63.6	80.2	67.2	75.6	113.7
<25 Min	75.1%	77.0%	67.3%	75.2%	70.3%	34.1%
<50 Min	85.5%	86.6%	80.8%	85.5%	82.6%	60.9%
<75 Min	91.5%	92.2%	88.8%	91.5%	89.9%	77.0%

Extra factors: Subtract 22.3 minutes of the FRT if the contractor has a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 11.1 minutes can be subtracted from the above table for a PGO-contract. Also subtract 23 minutes of the FRT if the failure is repaired during rush-hour.

Level Crossing Collision

Level Crossing Collision	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	112.4	102.8	71.4	117.7	106.6	82.8	67.9
85%	130.6	121.0	89.6	135.9	124.8	101.0	86.1
90%	155.7	146.1	114.7	161.0	149.9	126.1	111.2
<25 Min	28.7%	37.2%	59.6%	23.7%	33.9%	52.4%	61.6%
<50 Min	49.2%	55.7%	72.2%	45.3%	53.2%	66.9%	73.6%
<75 Min	64.6%	69.3%	81.1%	61.7%	67.5%	77.4%	82.1%

Level Crossing Collision	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	82.8	79.1	95.7	82.7	91.1	129.2
85%	101.0	97.3	113.9	100.9	109.3	147.4
90%	126.1	122.4	139.0	126.0	134.4	172.5
<25 Min	52.4%	54.8%	42.9%	52.4%	46.5%	12.3%
<50 Min	66.9%	68.7%	60.0%	67.0%	62.6%	36.0%
<75 Min	77.4%	78.7%	72.5%	77.4%	74.3%	54.8%

Extra factors: Subtract 23.1 minutes of the FRT if the failure is repaired during rush-hour.

Switch TOBS

Switch TOBS	
80%	82.9
85%	95.2
90%	112.0
<25 Min	35.2%
<50 Min	59.1%
<75 Min	76.1%

Extra factors: Subtract 20.7 minutes of the FRT if the contractor has a PGO-contract(in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 10.3 minutes can be subtracted from the above table for an PGO-contract.

Sign Out of Order

Sign Out of Order	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	100.1	49.0	81.8	77.4	38.4	59.1	45.9
85%	116.3	65.2	98.0	93.6	54.6	75.3	62.1
90%	139.2	88.1	120.9	116.5	77.5	98.2	85.0
<25 Min	24.2%	32.7%	45.2%	49.3%	74.6%	63.4%	71.0%
<50 Min	51.3%	56.8%	64.8%	67.5%	83.7%	76.5%	81.4%
<75 Min	68.8%	72.3%	77.4%	79.1%	89.6%	84.9%	88.1%

Sign Out of Order	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	110.1	105.1	62.0	29.4	95.1	75.7
85%	126.3	121.3	78.2	45.6	111.3	91.9
90%	149.2	144.2	101.1	68.5	134.2	114.8
<25 Min	9.5%	17.1%	61.4%	78.4%	30.6%	50.8%
<50 Min	41.9%	46.8%	75.3%	86.1%	55.5%	68.4%
<75 Min	62.7%	65.9%	84.1%	91.1%	71.4%	79.7%

Level Crossing Barrier not Responding Properly

Level Crossing Barrier not Responding Properly	
80%	28.8
85%	34.9
90%	43.0
<25 Min	76.2%
<50 Min	93.1%
<75 Min	98.3%

Extra factors: Add 28.6 minutes to the FRT if the failure is caused by maintenance activities.

Other Failures

Other Failures	
80%	106.9
85%	132.0
90%	167.9
<25 Min	46.8%
<50 Min	60.9%
<75 Min	71.0%

Extra factors: Subtract 78.7 minutes of the FRT if the failure is repaired during rush-hour.

Path Subsidence

Path Subsidence	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	189.9	156.9	184.9	169.9	161.9	101.9	103.9
85%	257.2	224.2	252.2	237.2	229.2	169.2	171.2
90%	288.9	255.9	283.9	268.9	260.9	200.9	202.9
<25 Min	5.6%	31.9%	9.4%	21.8%	28.1%	61.2%	60.5%
<50 Min	25.8%	47.7%	29.6%	40.1%	44.9%	69.2%	68.6%
<75 Min	43.2%	59.3%	46.1%	53.7%	57.3%	75.2%	74.7%

Path Subsidence	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	189.9	123.9	193.9	171.9	158.9	122.9
85%	257.2	191.2	261.2	239.2	226.2	190.2
90%	288.9	222.9	292.9	270.9	257.9	221.9
<25 Min	5.6%	51.8%	3.1%	20.1%	30.4%	52.3%
<50 Min	25.8%	62.3%	22.6%	38.8%	46.6%	62.7%
<75 Min	43.2%	70.0%	40.7%	52.8%	58.5%	70.3%

Status Path, Others

Status Path, Others	
80%	95.9
85%	116.6
90%	145.7
<25 Min	46.2%
<50 Min	62.1%
<75 Min	73.2%

Extra factors: Subtract 28.5 minutes to the FRT if the contractor is operating on a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 13.2 minutes can be subtracted from the above table for a PGO-contract. Add 95.3 minutes to the FRT if the failure is caused by maintenance activities.

Switch Disturbed

Switch Disturbed	
80%	85.7
85%	97.8
90%	113.7
<25 Min	35.6%
<50 Min	57.7%
<75 Min	74.5%

Extra factors: subtract 20.7 minutes of the FRT if the contractor is operating on a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 10.3 minutes can be subtracted from the above table for a PGO-contract.

Level Crossing Specific Damage

Level Crossing Specific Damage	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	71.0	80.6	91.0	65.7	65.2	44.1	26.5
85%	86.5	96.1	106.5	81.2	80.7	59.6	42.0
90%	108.4	118.0	128.4	103.1	102.6	81.5	63.9
<25 Min	52.3%	42.6%	29.6%	56.9%	57.3%	71.4%	79.4%
<50 Min	70.4%	64.5%	56.7%	73.2%	73.4%	82.1%	87.1%
<75 Min	81.4%	77.8%	73.0%	83.2%	83.3%	88.7%	91.8%

Level Crossing Specific Damage	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	41.4	37.7	54.3	41.3	49.7	87.8
85%	56.9	53.2	69.8	56.8	65.2	103.3
90%	78.8	75.1	91.7	78.7	87.1	125.2
<25 Min	72.8%	74.6%	65.3%	72.9%	68.2%	33.9%
<50 Min	83.0%	84.1%	78.3%	83.0%	80.1%	59.2%
<75 Min	89.3%	90.0%	86.4%	89.3%	87.5%	74.6%

Extra factors: Subtract 22.3 minutes of the FRT if the contractor has a PGO-contract, also subtract 23 minutes of the FRT if the failure is repaired during rush-hour.

Buckling in Path

Buckling in Path	
80%	101.1
85%	140.8
90%	156.8
<25 Min	37.7%
<50 Min	58.1%
<75 Min	71.2%

Extra factors: Subtract 28.5 minutes to the FRT if the contractor is operating on a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 13.2 minutes can be subtracted from the above table for a PGO-contract. Add 95.3 minutes to the FRT if the failure is caused by maintenance activities.

Sign Turned Off

Sign Turned Off	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	41.0	15.5	31.9	29.7	10.2	20.5	13.9
85%	47.0	21.5	37.9	35.7	16.2	26.5	19.9
90%	54.6	29.1	45.5	43.3	23.8	34.1	27.5
<25 Min	62.7%	87.5%	73.2%	75.5%	90.7%	83.8%	88.5%
<50 Min	87.1%	97.5%	92.3%	93.3%	98.3%	96.4%	97.8%
<75 Min	97.4%	99.7%	98.7%	99.0%	99.8%	99.5%	99.8%

Sign Turned Off	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	51.0	49.7	22.0	5.7	38.5	28.8
85%	57.0	55.7	28.0	11.7	44.5	34.8
90%	64.6	63.3	35.6	19.3	52.1	42.4
<25 Min	50.0%	51.7%	82.6%	92.9%	65.7%	76.4%
<50 Min	79.1%	80.3%	96.0%	98.9%	88.7%	93.7%
<75 Min	94.7%	95.2%	99.5%	99.9%	97.8%	99.0%

Smouldering Sleepers

Smouldering Sleepers	
80%	8.9
85%	15.4
90%	18.7
<25 Min	93.4%
<50 Min	98.5%
<75 Min	99.6%

Extra factors: Subtract 28.5 minutes to the FRT if the contractor is operating on a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 13.2 minutes can be subtracted from the above table for a PGO-contract. Add 95.3 minutes to the FRT if the failure is caused by maintenance activities.

Level Crossing TOBS

Level Crossing TOBS	Arnhem	Amersfoort	Alkmaar	Amsterdam	Eindhoven	Groningen	The Hague
80%	92.7	102.3	112.7	87.4	86.9	65.8	48.2
85%	115.0	124.6	135.0	109.7	109.2	88.1	70.5
90%	147.3	156.9	167.3	142.0	141.5	120.4	102.8
<25 Min	48.3%	39.9%	28.6%	52.3%	52.7%	65.2%	72.7%
<50 Min	64.2%	58.9%	52.1%	66.8%	67.0%	75.3%	80.5%
<75 Min	74.7%	71.2%	66.7%	76.4%	76.6%	82.3%	85.8%

Level Crossing TOBS	Kijfhoek	Maastricht	Roosendaal	Rotterdam	Utrecht	Zwolle
80%	63.1	59.4	76.0	63.0	71.4	109.5
85%	85.4	81.7	98.3	85.3	93.7	131.8
90%	117.7	114.0	130.6	117.6	126.0	164.1
<25 Min	66.5%	68.2%	59.7%	66.5%	62.3%	32.4%
<50 Min	76.2%	77.4%	71.7%	76.2%	73.4%	54.3%
<75 Min	82.9%	83.7%	79.7%	82.9%	80.9%	68.1%

Bridge Disturbed

Bridge Disturbed	
80%	56.9
85%	70.5
90%	89.7
<25 Min	60.9%
<50 Min	76.9%
<75 Min	89.4%

Extra factors: Subtract 28.5 minutes to the FRT if the contractor is operating on a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 13.2 minutes can be subtracted from the above table for a PGO-contract. Add 95.3 minutes to the FRT if the failure is caused by maintenance activities.

Overpass Collision

Overpass Collision	
80%	2.8
85%	6.5
90%	11.5
<25 Min	96.8%
<50 Min	99.7%
<75 Min	99.8%

Extra factors: Subtract 28.5 minutes to the FRT if the contractor is operating on a PGO-contract (in comparison with the FRT of the OPC-contract). Since both contracts are present in the same amount, 13.2 minutes can be subtracted from the above table for a PGO-contract. Add 95.3 minutes to the FRT if the failure is caused by maintenance activities.

Appendix H: Influence of District Control Office

H.1. Influence of the PCA

The PCA (the term used at ProRail for the contractor) takes care of the actual repair of the failure. Thus, when a switch can no longer be controlled, he is the one wielding the welder to repair it. ProRail merely informs them about the fact that a failure is occurring and oversees the entire chain of events following this.

Every infra-object is taken care of by exactly one of the four PCAs (Asset-rail, BAM Rail, Strukton and VolkerRail). However, it is possible that multiple PCAs operate in one district control office area. In, for instance, district control office area Maastricht only Strukton Rail is active but all of the PCAs are active in the district control office area Utrecht. Which PCA operates where depends upon who is selected for which contract of an infrastructure area.

Since every contractor has a specific way of operating, different levels of specialisation and different goals, their performance tends to differ. Which PCA operates on a certain infra-equipment is therefore a possible important variable for the prediction of the failure recovery time (FRT).

First, whether or not there is a significant difference between the PCAs is determined. This is done by executing t-tests over all the FRTs per PCA comparison. As explained in section 4.2, a t-test (or student-test) can be used to look for a significant difference between two data sets. Again, a significance level of 0.05 is used. When the p-value of a comparison is higher than this significance level, there is no reason to assume a difference between the two data sets. When the p-value is lower than the level, there is enough reason to assume a difference (Craparo, 2007). The highest five percent of the FRTs are omitted from this research since these influence the test-results too much.

The p-values for the comparisons between the PCAs are shown in Table H.1. As can be seen from the table, the FRTs of VolkerRail clearly differ from the FRTs of other PCAs. However, while Strukton and BAM Rail also differ, the position of Asset-Rail is not clear: it doesn't significantly differ from both of these two PCAs. Since the p-value of the comparison between Asset-Rail and Strukton is the highest, these two are the closest to each other. Therefore, the data of these two is combined. When a new t-test comparing the FRTs of BAM-Rail with the FRTs of Asset-Rail and Strukton is run a p-value of 0.0340 is found. The combination of these PCAs is thus significantly different from BAM-Rail. In the reminder of this section the distinction between the PCAs: Asset-Rail + Strukton, BAM Rail and VolkerRail is made.

Table H.1: The T-Test Values per PCA-Comparison.

Comparison PCAs	P-value
Asset-Rail – BAM Rail	0.287
Asset-Rail – Strukton	0.498
Asset Rail - VolkerRail	0.000171
BAM Rail – Strukton	0.0265
BAM Rail – VolkerRail	0.000259
Strukton - VolkerRail	0.00000564

There is thus a significant difference between these three PCA-groups. However, in order to predict the FRT, the difference of the FRT between the PCAs per failure form has to be researched and quantified. Therefore, t-tests are used to look for differences between the FRT between PCAs on certain failure forms. The failure forms representing the general categories: path (status and TOBS), switch, level crossing and sign are used. These represent the other failure forms.

An overview of the results of these t-tests is shown in Table H.2.

Table H.2: The T-Tests comparing the FRTs of Different PCAs on the Most Important Failure Forms.

	Asset-Rail/Strukton – BAM Rail	Asset-Rail/Strukton – VolkerRail	BAM-Rail - VolkerRail
TOBS Path	0.13	**0.0014**	0.21
Switch NIC	0.63	**0.000019**	**0.0000055**
Level Crossing Failure	0.45	0.27	**0.09**
Level Crossing Collision	0.51	**0.0010**	**0.077**
TOBS Switch	0.27	0.46	0.60
Sign Out of Order	**0.0012**	**0.014**	**0.0000046**
Status Path	0.154	**0.072**	0.45

The results of the Table H.2 can be quantified by looking at the difference in average per PCA for the failure forms of the comparisons with a significant difference, thus, with a p-value below 0.1 (0.1 is used now since the amount of data per failure form is a bit lower than without the failure forms).

The quantified results are given in Table H.3. It is important to note that, when there is one PCA performing differently and the others the same these are grouped into one. This is indicated with the dash. Per failure form, the average FRT in minutes is given for the PCA with the lowest FRT while the difference in minutes is shown with the numbers with the plus symbol in front of it. Thus, while the average FRT for the failure form TOBS path for PCA Asset-Rail/Strukton is 56 minutes, this is 17 minutes higher for the PCA BAM Rail (73) and 34 minutes higher for VolkerRail (90).

Table H.3: The Average FRT (Minutes) of the Best Performing PCA and the Difference of the other PCAs per Failure Form.

	Asset-Rail/Strukton	BAM Rail	VolkerRail
TOBS Path	56	+17	+34
Switch NIC	35	-	+27
Level Crossing Failure	+5	27	+12
Level Crossing Collision	50	-	+34
TOBS Switch	36	-	-
Sign Out of Order	+14	24	+22
Status Path	+22	+5	55

H.2. Influence of the Movability around the Path

The movability around the path fluctuates a lot at different route sections ("baanvakken"). Some sections of the path continuously have a road lying next to it. When the mechanic has to go to the failure location or needs different locations to diagnose the failure, he can use the car to immediately get to the right location. However, there are also multiple route sections which do not have such a road. Moving around in such a route section requires access points, like level crossings, or a road running relatively close to the path. Depending on the situation, this can result in multiple kilometres to be walked by the mechanic. This obviously does not improve the FRT.

Since there are multiple positive and negative access points per district control office this can't be generalized per district control office. Therefore, individual route sections are compared based on their movability score. These scores are the scores mentioned by the district control office when confronted with this theory. Examples of well or badly accessible and movability of the mechanic route sections are:

- *Schagen-Heerhugowaard.* District control office Alkmaar, movability score 2. Continuously surrounded by ditches and meadows and barely any level crossings makes specific locations on this route sections very hard to reach. Average failure recovery time for a path TOBS: 185.4 minutes.
- *Uitgeest to surrounding cities.* District control office Alkmaar, movability score 8. There are multiple level crossings close to these route sections. Besides this, there are a lot of roads running next to the tracks with multiple parking spots. Average failure recovery time for a path TOBS: 150.4 minutes.
- *Sittard – Roermond.* District control office Maastricht, movability score 8. This area is easily accessible due to the road running next to the track. Average failure recovery time for a path TOBS: 56.6 minutes.
- *Sittard – Maastricht.* District control office Maastricht, movability score 3. This is a mountainous area and therefore hardly accessible by car. Average failure recovery time for a path TOBS: 76.3 minutes.

Since these route sections are specifically mentioned by the district control offices, the assumption is made that these are the extreme cases. Thus, the route sections not mentioned would have an average movability score of 5.5. When the mentioned route sections are ignored, the average FRT for a path TOBS at district control office Maastricht is 65.2 and at Alkmaar this is 162.2.

Till now, the district control offices gave very specific scores per route section (2, 3 and 8), however, the offices noted that it is very difficult for them to give such a specific score while being accurate. Therefore, scoring the movability of a route section will be simplified by allowing only three scores: low, medium and good. The previous scores below 4 are now classified as low while a score above 6 is classified as good.

To see whether or not these differences are significant, again, t-tests are used. The results of the t-tests comparing the different levels are the following:

- T-test, low-high score: 0.0084
- T-test, low-medium score: 0.042
- T-test, medium-high score: 0.030

Thus, the t-tests showed that the differences between the different movability scores are significant. The factor: movability around the path is therefore relevant for the research.

Quantifying the factor can again be done by comparing the average FRT per score to each other. This is done for the same failure form and the same district control office. This shows that a high movability score results in a ten minute lower FRT while a low accessibility score increases the FRT with twelve minutes.

H.3. Influence of a Sharp Corner in the Path

A train usually barely has to take any extremely sharp corners. However, there are circumstances when there isn't enough space to create an infrastructure with only moderate corners, usually in cities. Due to this, a sharp corner has to be placed. A sharp corner will increase the friction of the wheels of the train on the outside rails. Due to this, more metal particles will be created, blocking the ES-joints. Since these failures are easy to repair, a path TOBS in sharp corner will have a low FRT.

Multiple district control offices verified that a sharp corner indeed resulted in more metal particles responsible for the path TOBS. Since this effect is the same as the grinding train, we can conclude that a sharp corner decreases the expected FRT of a path TOBS with 30 minutes (see chapter 4 and 5).

H.4. Influence of the Age of the Infrastructure

The infrastructure of the railway typically is relatively old. Rails of more than 30 years old aren't uncommon in the train industry. However, when a rail reaches a certain age and, consequently, is used a lot unequally wear will cause a high amount metal particles and, thus, TOBS due to metal particles on the ES-joint. Whether or not this has a significant influence is researched in this section.

The distinction has to be made, according to the district control offices, between old rails and normal rails. There is no need for more categories due to the fact that a new or medium used rail both don't result in an excessive amount of metal particles, only too old rails do. Since this, again, is only one factor which is either there or not (the rails is old or not) quantifying it is relatively easy: the average when the rails are old is compared with the average when the rails are not old. This results in a difference of fourteen minutes when the rails are old in comparison with normal rails. This is based upon a route section in The Hague. The district control office noted that this section is really old, it therefore resulted in 80 percent of its path TOBS resulting from metal particles.

H.5. Influence of the Distance of Spare Parts

According to the district control office, this directly follows from the PCA active in a certain region. When a PCA decides to allocate more spare parts to their regions depot, the FRT used to get a spare part will go down. Since stock costs money, one PCA can decide to stock more spare parts in their depot than the other. Thus the spare part recovery time is a consequence of the PCA active in a certain region and will no longer be included in this analysis.

H.6. Influence of the Continuity of the Same PCA

Every five years a contractor (PCA) is allowed to offer a bid for the contract of a certain region. When this bid is accepted, this PCA is responsible for the maintenance of this region. Thus, it is possible that the maintenance activities done is certain region are not done by the same PCA that did this last year. A switch of PCA in a region has a negative impact on the expected failure recovery time (FRT). This is caused by several reasons.

The first is the unfamiliarity with the failed infra-object. The failure history and the best failure recovery method are not known, causing the FRT to be higher. Another reason for the increased FRT is the unfamiliarity of the mechanic with the area and its traffic opportunities. Thus, a mechanic does not know the efficient method to get to a certain location causing an increased travelling time.

This could be solved by the information sharing of different PCAs, however, due to rivalries and EU-regulation, this is not done/allowed. Because of this, it takes several years for a PCA to get used to the most efficient and quickest way to solve a certain failure in the region.

According to all the district control offices visited (The Hague, Alkmaar, Utrecht, Maastricht, Zwolle and Groningen) a switch of PCA has a significant effect on the FRT. However, since the bids for contracts are once in five years and the next one is this year (2016), the last contract change is a long time ago. Unfortunately, the data necessary for this research is no longer available. Thus, while this is a significant factor, it can only occur once in five years and only for an area which switched PCA. However, the quantified effect can't be determined.

Appendix I: Temporary Failure Recovery

In this chapter another type of failure recovery is analysed: temporary failure recovery. While the focus in chapters in this research is on the full recovery of a failure, now the focus is on the temporary failure recovery. Sometimes it is better to execute a temporary failure recovery while the full recovery will come later during a more suitable time. This chapter will first explore differences between the different failure recovery types. After this, the performance of temporary failure recovery is explored and compared with full recovery. The goal of the chapter is to help VL (traffic control) decide whether to let the mechanic execute a temporary solution or to execute a full repair. This is another set of possibilities for the decision tool about the FRM.

I.1. Difference Full and Temporary Failure Recovery

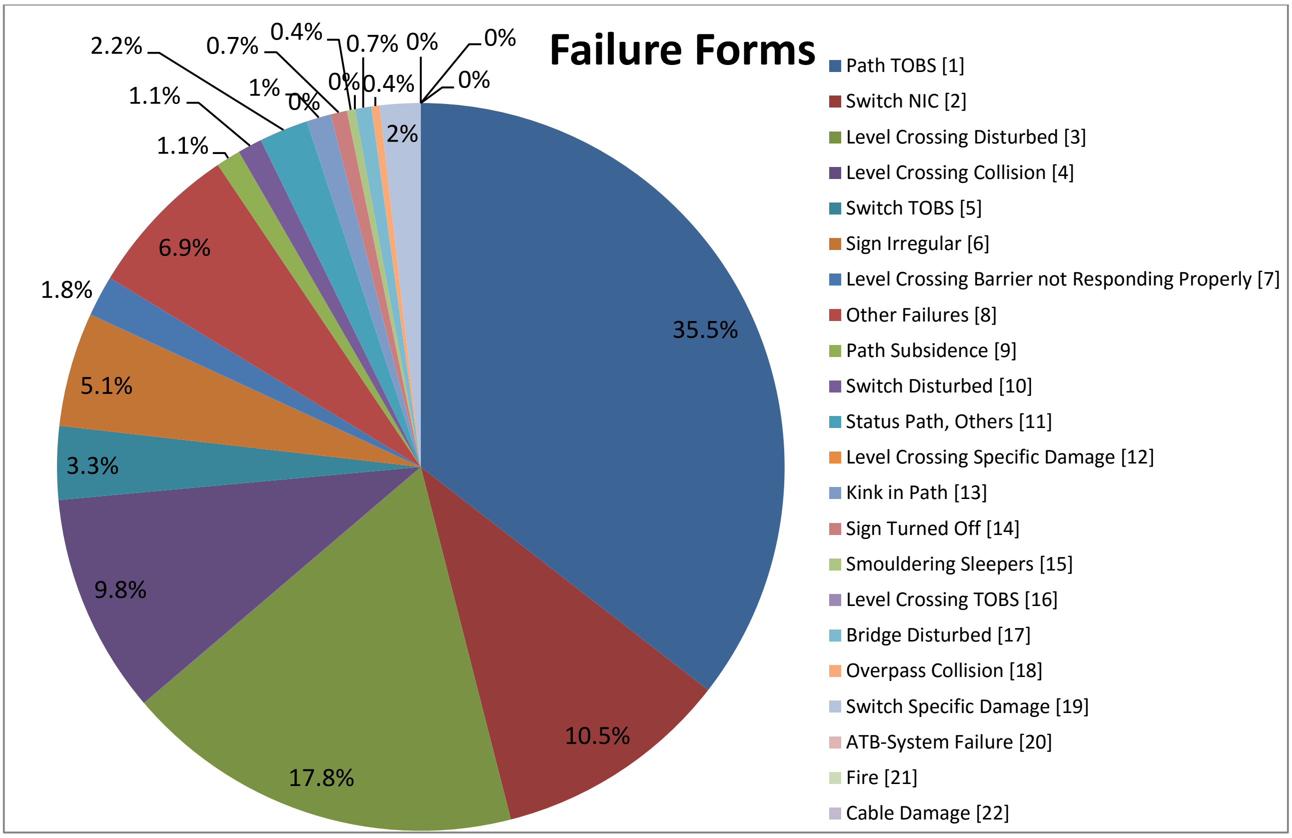
The failure recovery at ProRail consists out of multiple types. This type indicates the degree of recovery of the functionality of the infra. ProRail makes use of four different failure recovery types displayed in Table I.1. As can be seen from the table, a full failure recovery is by far the most occurring failure recovery type, followed by temporary failure recovery.

Table I.1: The Types of Failure Recovery applied on Urgent Failures and their Occurrence.

Type Failure Recovery	Work done by Contractor	Example	Status Infra-Object	Occurrence
Full Recovery	The contractor finished the failure recovery process	Level crossing barrier (“overweg-boom”) replaced.	Object functions for 100%	71.2%
Temporary Recovery	The contractor used a temporary solution to make the object function again. He has to come back at a later date.	Level crossing barrier of another type placed (only spare part available). Correct barrier is placed later.	Object functions for 100%	16.5%
Partial Recovery	The contractor used a failure recovery method which only partially solves the problem. He has to come back at a later date.	Switch fixed in place. Only one direction of path is usable.	Object functions partially	2.7%
No Recovery	The contractor did his research but not any failure recovery work.	As soon as the mechanic arrived at the failed switch, the switch started working again. It still worked after half an hour thus the mechanic left.	Object functions for 100% / Object does not function	9.6%

Since a temporary failure recovery is executed for different reasons and under different circumstances than a full failure recovery, a different distribution of failure occurrences over the failure forms can be expected. As shown by Figure I.1, this is indeed the case. The failure forms are plotted in the pie chart in the same order as they were plotted as in the pie chart of Figure 4.1, which is in the order of frequent occurrence when a full recovery is executed. The figure shows that a temporary failure recovery tends to be especially more frequent for path TOBS and level crossing failures.

Figure I.1: The Percentage of Temporary Failure Recoveries of a Specific Failure Form as compared to the Total Amount of Occurrences.



This increase in frequency can be explained by a look into the motivation for temporary failure recovery instead of a full recovery. A common situation when temporary failure recovery takes place is when the spare part is not present but the failure has to be repaired. Instead of waiting a long time for the spare part to be recovered, another, less suitable, part is used (like in the example in Table I.1). Another typical temporary solution is the placement of a "slipkabel". This is a cable which is placed to replace the energy supply of another, possibly failed, cable. It is often not clear where in this cable the failure is. It is also possible that, if the cable failure is located, repairing the failure takes too much time. For these cases, it can be more attractive to place a temporary, replacing slipkabel to get the energy supply up and working again.

When analysing the temporary failure data, three types of motivations to execute a temporary instead of a full failure recovery are found. Each of these results in a different kind of question to be asked:

- 75% of the temporary failure recoveries are executed because of the faster solution they offer to the problem. Thus, the problem is known and the mechanic makes, in consultation with the Trdl, the decision to execute a faster, temporary failure recovery based on the expected repair time and the necessity to use the path at the current moment. It is important to compare the repair times (thus, without the diagnosis time) of a temporary failure recovery with a full failure recovery in this case. This helps the mechanic and the Trdl to make the right decision about the type of failure recovery to execute.

- 19% is executed because there is no spare part present. Like in the example of Table I.1, a different type of spare part or even the old, partially repaired, part is used. The only way to directly execute a full failure recovery would be to wait for the spare part to arrive. In this case, whether or not this waiting time is worth it is the most important consideration.
- The last 6% of the temporary failure recoveries are executed because the direct cause of the failure could not be found, but this temporary solution would surely solve it for now (like a long slipkabel when the exact failure location in the cable has not been found yet). The question to be asked here is whether or not to spend more time to search for the exact cause. However, since this motivation only represents a low amount of the failures and the marginal value added when looking longer is expected to be low, the assumption is made that the mechanic and the Trdl make the right call in this case.

I.2. Temporary Failure Recovery Decision – Time Advantage

When the decision for a temporary failure recovery instead of a full failure recovery is made knowingly, there has to be a significant advantage in expected repair time between the two. Especially since a temporary failure recovery does not permanently fix a failure but is followed by a full failure recovery on a more suitable time.

To determine the repair time, the previously used time in chapter four has to be split up to extract the repair time from the combined time of diagnosis and repair. The actual start of repair is not recorded, therefore the time the second prognosis is given is used as the start of the repair. A mechanic is required to give a prognosis to the Trdl about the time he expects the failure to be repaired: after the initial prognosis, this is the second prognosis. The assumption is made that this is the time when he knows enough about the failure to finish diagnosing and start the actual repair. This is largely confirmed after conversations with the mechanics and by the afterwards explanation about a failure given in the database.

A problem which has to be tackled first is the amount of wrong data and because of this: the bias. 70 Percent of the failure recovery prognosis given by the mechanic are given after the failure is repaired. This makes it obviously easier for the mechanic to predict the correct failure recovery time but makes 70 percent of the data unusable in the repair time determination. However, the main problem comes from the prognosis which are done before the failure recovery. Because a mechanic is not motivated to give a prognosis beforehand when only a simple repair is necessary, the remaining 30 percent of the data represents more of the long FRTs than the short FRTs. When data is not missing completely at random, a bias is introduced (Ibrahim, Chen, Lipsitz, & Herring, 2005). In this case, towards a longer repair time.

This problem is dealt with by weighting a repair time by the amount of data points it represents. For instance: when three out of hundred repair times (which are correctly registered) represent the lowest ten percent of all the failure recovery data, they get a weight of ten percent instead of three percent (3/100).This does not completely eliminate the bias but, with sufficient data, removes the largest negative effects. This method, however, is not suitable when the probability distribution has to be determined since ten percent intervals group together too much of the individual data points for an accurate probability distribution estimate. Fortunately, only the absolute difference between the repair time of a full and a temporary failure recovery is necessary. Thus, this is not necessary. The average repair time suffices for this.

To give a reliable conclusion, the averages are only determined for the four most occurring failure forms: path TOBS, switch NIC, level crossing disturbed and level crossing collision. These represent almost three quarters of all the data and have sufficient data points to draw a reliable conclusion. Based on the previously explained method the average repair time was determined and displayed in Table I.2

Table I.2: The Average Repair Times of Full and Temporary Failure Recoveries of the Four Most Occurring Failure Forms.

	Full Failure Recovery	Temporary Failure Recovery
Path TOBS	76.5 Minutes*	43.5 Minutes*
Switch NIC	64.1 Minutes*	34.9 Minutes*
Level Crossing Disturbed	65.1 Minutes*	35.0 Minutes*
Level Crossing Collision	-**	26.7 Minutes**

* Placing a slipkabel takes 119 minutes on average. These are not included in the average calculations.
**Replacing a level crossing barrier (with a new or old barrier) takes 110 minutes on average. These times are not included in the above average of the failure form: level crossing collision. Since a full failure recovery with a level crossing collision always requires a barrier to be replaced (instead of, for instance, using tape as a temporary solution before replacing the barrier) there are no other times present for this full failure recovery form.

In Table I.2, two specific large failure recovery methods are omitted: the placement of a slipkabel and replacing a level crossing barrier. It is easy to recognize early in the failure recovery process that this is necessary since both of them have a large failure recovery time. This leaves the normal failures.

The table shows that for the normal failures a temporary failure recovery is about half an hour faster than a full failure recovery. Thus, it requires more than one train free period (25 minutes) less on average. However, when a temporary failure recovery is executed, the full failure recovery still has to be executed but now, during a more suitable time (this is usually the next night) whether or not this thirty minutes less repair time during the day is worth the approximately 65 minutes more repair time during the night depends on the situation. VL has to make this decision based on the factors relevant for the failure recovery moment (identified during interviews with Trdls and DVLs): the severity of the path withdrawal at that location, the intensity of train-traffic at the time and the severity of the failure recovery. The latter factor is clear now, however, the former two are based on the specific situation and should therefore be taken into consideration by the expert himself (the Trdl and DVL responsible for the paths of the failure location)

I.3. Temporary Failure Recovery Decision –Absence of Spare Part

Every one in five temporary failure recoveries are executed because there is no spare part present. A mechanic always brings a fixed set of spare parts with him to a job with an unknown cause. When the cause, and thus: the required action, is known, he can bring the spare parts which are necessary for that specific job. But when the cause is not known, he hopes the spare part is in the set of spare parts he brought. Occasionally, this is not the case. This happens often when the part which has to be replaced is large, expensive and a failure of it uncommon, these spare parts are therefore not brought in the fixed set. A common example of this is the level crossing barrier failure explained above. Since there are multiple types of large level crossing barriers used at a level crossing it is not preferable when a mechanic takes all of these barriers with him for every failure. When a barrier has to be replaced, he therefore rarely has the right barrier with him. This correct barrier has first to be picked up from the local (the contractors depot present at every district control office region) or even central depot (every contractor has one or a few central depots in The Netherlands).

Relatively often, a mechanic choses to execute a temporary failure recovery first. Therefore, the duration of getting the spare part has to be compared with the expected repair time of a temporary failure recovery.

The latter was already determined in the previous section (see Table I.2) while the spare part recovery time can be calculated from the text columns in the data files. These columns show that the average spare part recovery takes 150 minutes. However, if a spare part has to come from the supplier, instead of the local or central depot, it usually takes way longer. Therefore, it is a good idea to execute a temporary solution when a spare part has to come from the supplier since the delivery time takes too long (unless the supplier explains beforehand that he is capable of delivering it earlier, which is a specific case). When the spare part comes from the central depot, 150 minutes is used as an indication.

When a level crossing barrier has to be replaced, the expected time to do this is, as indicated above, 110 minutes. Thus, when a temporary barrier is placed while a new barrier is getting delivered from the depot, the temporary barrier can't even function for an hour before the correct barrier is present. Thus, 110 minutes downtime for less than one hour of functioning time. It is therefore advisable to arrange traffic agents or let the mechanics perform as traffic agents while someone is getting the correct barrier, if possible. By doing this, the level crossing can still be fully used before the correct part arrives.

If the problem is another failure instead of a level crossing barrier to be replaced, the expected temporary failure recovery repair time is way lower (25-35 minutes) and can often be repaired in one or two 25-minute periods. This is way lower than the time the equipment can function before the correct spare part arrives. Besides this, it is often not possible to use traffic agents or other solutions to keep the equipment functioning. In this case, it is therefore advised to execute the temporary failure recovery. However, the expected length of the delivery process in that specific situation should be taken into account.

I.3. Conclusion

During this chapter the difference between full and temporary function recovery and its failure recovery length was researched. While a full failure recovery repairs a failure permanently, a temporary failure recovery is faster but only lasts for a certain amount of time. However, the permanent repair which follows the temporary repair can be done during a more suitable time.

Three types of temporary failure recovery were identified based on their motivation: recovery motivated by the lower repair time, a lack of spare parts and an insufficient prognosis. While the first represents three out of four cases, the last one only occurs occasionally and can therefore be ignored.

When a temporary failure recovery is executed because of the lower repair times in comparison with a full failure recovery it is, on average, half an hour faster (besides when a slipkabel or a different level crossing barrier will be placed). This is independent of the failure form. However, due to the influence of situation-specific factors in the form of the severity of the path withdrawal and the current intensity of the train traffic it is not possible to give a conclusion about which of the two recovery types is more suitable in which case. Whether or not to use this extra half an hour quicker repair at the cost of a full repair at a later period should be determined by the expert based on the current situation-specific elements.

When a temporary failure recovery is executed because of the lack of spare parts, another consideration has to be done: whether or not the temporary repair time is low enough in comparison with the time it takes to get a new part.

Placing a different type of level crossing barrier as a temporary solution while the correct level crossing barrier is being delivered is often not worth the trouble. The delivery is often present within an hour after placing the other barrier. It is therefore better to use another solution, for example: traffic agents, which keeps the level crossing functioning longer. This will hinder the train traffic far less.

In most cases however, the temporary repair time is far lower than the delivery time of the spare part. In those cases, it is worth to first execute the temporary solution before waiting for the spare part. Due to the high variability in the delivery time, it is important to quickly try to determine this time. In some cases the spare part depot is close while in other cases, it should come from a supplier which is situated very far. Taking this variability into account is important for the VL when determining when to repair a failure.

Appendix J: The Evaluation of the Choices of the Traffic Control about the Function Recovery Moment

When an urgent failure occurs, traffic control can, in general, make two choices: repair the failure now or repair it during the night. Which choice to make depends on several factors: the severity of the failure, the current business of train traffic and multiple other factors. In this appendix, a method is developed to evaluate the choice the traffic control has made. By evaluating this choice, traffic control is motivated to think more deeply about the factors influencing the decision next time. This appendix should thus be used as an addition to the failure recovery moment (FRM) tool of chapter 7.

J.1. Relevancy of the Problem

When a certain failure occurs it is possible for the traffic control (VL) to either repair the failure direct (priority 2) or to postpone the failure recovery till later that day or night by arranging a certain repair start time with the contractor(priority 5). Usually, a repair with an immediate, serious effect on the train traffic is repaired directly while a repair of a failure with a less serious effect is postponed. However, in some cases, this decision is hard to make and, being human, traffic controllers occasionally make the wrong, possibly biased, decision in those cases. In such a case, the train traffic is hindered more than necessary.

The criterion used by the OBI and Trdl to decide whether to directly repair a failure or postpone is the availability of direct access to the infrastructure. When this is, in the opinion of the operators, possible, the failure is repaired directly. When this is not possible, the failure is repaired later based on a time appointment. Deciding whether direct access to the infrastructure is possible or not is often difficult.

An example of this is the situation where a switch can't be controlled (NIC) at 10:30 in the morning. A Trdl (movement inspector) makes, in consultation with the OBI, the call to repair it in the night (from 01:00 till 04:00 the mechanic gets access to the paths). However, after having caused sixteen trains to use a different route with a corresponding delay of fifteen minutes, the failure is repaired that night in only half an hour, this could have been done during the day by only delaying four trains half an hour!

Now, a Trdl is not familiar with the exact effect of his decision. However, by letting the Trdl reflect on his/her decision he/she is capable of making a more conscious decision about the failure recovery moment (FRM). This increases the probability of making the best decision when the same situation occurs the next time.

J.2. Necessary Information for the Decision-Making Process

To be able to evaluate the decision made, the advantages and disadvantages of the alternatives have to be clear. With the help of experts on the subject, these are determined and shown in Table J.1.

The most important advantage of a repair in the night is the favourable time when the repair occurs. However, the price to be paid is to deal with a none- to partially-functioning infra-equipment till the evening. A danger of postponing the repair action till the night is the lack of diagnosis done. The severity of the failure and the necessary tools to repair the failure are therefore not known. When this diagnosis is done during the day, a more precise prognosis about the failure recovery duration can already be done and a more educated decision about the

FRM can be made. This lowers the risk of unexpectedly having to execute a large repair operation.

Table J.1: The Advantages and Disadvantages of the Repair Time Alternatives.

	Alternative 1: Repair Now	Alternative 2: Repair Later
Advantages	-Infra-equipment can fully be used the rest of the day. -Quick diagnosis of the extent of the failure. -Possibility to move the repair to the evening if diagnosis is unfavourable.	-Path withdrawal happens at a suitable time. - No repair interruptions from the train traffic.
Disadvantages	-Path withdrawal at a relatively bad time. -Occasional repair interruption due to necessary train traffic.	-Infra-equipment does not fully function for the coming X hours. -Unpredictable failure recovery time due to lack of diagnosis.

From Table J.1 the four most important factors influencing the failure recovery moment decision can be seen:

- Train traffic severity at the nearest possible repair moment and hindrance for this traffic based on the failure recovery time (FRT);
- Time and train traffic severity till the night and the continuous hindrance for this if the failure is not repaired now;
- Costs for a repair during the night based on the FRT of the failure;
- Risk for insufficient time during the night to execute the repair activities due to an unexpected severity of the failure.

J.3. Deming Circle

The Deming Circle is a management method often referred to as Plan-Do-Check- Act (PCDA). It is used for process control and the continuous improvement of processes. The circle makes use of four steps, plan, do, check and act, to monitor and control a process and, by iterating the steps, improving it (Rother, 2010). The Deming Circle is illustrated in Figure J.1. As can be seen from the figure, it is important to continue applying the circle after one iteration. By doing this, the actions in the next cycle can be improvements to the actions in the last cycle. Thereby, the process is improved.

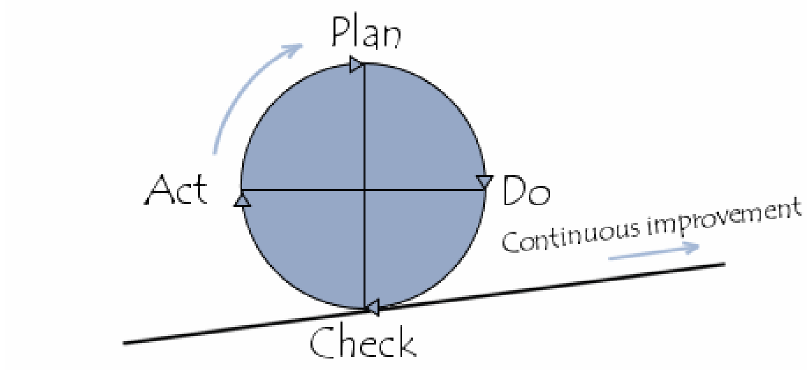


Figure J.1: The Deming Circle of Continuous Improvement.

The Deming Circle is a tool used more often at ProRail and will therefore be applied on the decision process described above. This is done by describing each of the four steps of the Deming Circle. First, the theoretical meaning of each step is described. This is followed by the application of the step on the process of ProRail.

Plan: During this step the goals of the improvement of the process are set. This is followed by the method/planning used to achieve these goals. In case of the priority decision process at ProRail the goal is to select the priority level: two (repair as quickly as possible) or five (urgent repair with a time appointment which is, often, during the night) which results in the least amount of delayed passengers. The most difficult part of this step is the way to make the best decision.

Section J.2 identified the factors influencing this decision. Combining this with the information about the failure recovery process given in Chapter two, the two decision to be made can be compared. These are compared on their influence on the goal of the decision making process: minimal delay for passengers. The total delay per alternative can be determined by the following formulas:

Alternative 1 – Repair now: Total delay = Expected number of passengers having to make use of the unavailable infra-equipment's due to failure recovery per hour when failure recovery occurs * Expected FRT * Expected number of minutes delay per passenger due to the unavailability of infra-equipment's.

Alternative 2 – Repair later: Total delay = Expected number of passengers having to make use of the failed infra-equipment per hour* Time remaining till evening * Expected number of minutes delay per passenger due to the unavailability of the infra-equipment.

The most important variables per alternative compared to each other are shown in table J.2.

Table J.2: The most important Variables influencing the Total Delay per Alternative compared to their Counterparts per Alternative.

Alternative 1: Repair Now	Alternative 2: Repair Later
Number of passenger influenced per hour	**Number of Passengers Influenced per hour**
Failure Recovery Time	**Time till Evening**
Minutes Delay per Passenger **+**	**Minutes Delay per Passenger** **+**
Total Delay	**Total Delay**

By looking at the formulas, the table and the situation the major differences between the situations can be seen: the total time on a day a failure influences passengers and the numbers of passengers delayed due to the failure. In case of a direct repair, the total time on a day a failure influences the passengers is relatively low (only the FRT) while the number of passengers delayed per hour is relatively high (due to multiple paths having to be withdrawn when repairing the failure). For the second alternative, this is the other way around: the total time on a day a failure influences the passengers is high (the time till the evening) while the number of passengers influenced per hour is low (cause only one path has problems).

Thus, four variables, two pairs, mainly influence the decision. The following variables have to be weighed against each other by the Trdl.

Failure recovery time	⇔	*Time till evening.*
Number of passengers influenced per hour by the WBI	⇔	*Number of passenger influenced per hour by the failure.*

*A WBI ("Werkplek Beveiliging Instructie" or workplace safety instructions) is a description of the actions necessary for maintenance to the rail-infrastructure per infra-equipment. Since this also comprises the train paths to be blocked when a certain object is repaired, the WBI should be used when this decision is made.

The Trdl makes the decision which failure recovery moment to use and he executes it in the next step.

Do: Like explained in the step above, the decision made is implemented and executed in this step. Also, the measurements are taken which are necessary to check whether the goal is reached. Thus, the actual performance of the actions taken is measured by looking at the amount of delayed passengers. Therefore, the values of the before mentioned variables, depending on the choice made, are checked.

Check: In this step the results of the decision taken are compared against the goals set in the planning step. Since the goal was to select the priority level resulting in the least amount of delayed passengers the actual amount of delayed passengers has to be compared to the expected amount. A distinction between the two decisions/priority levels has to be made in this step. Per priority level, the actual value of the variables shown in Table J.2 should be determined. For every priority level this comes down to the next questions having to be asked:

- Priority level 2.
    - Is the failure recovery time (FRT) as expected?
    - Is the number of hindered passengers per hour as expected?
    - Other unexpected negative side-effect of the actions taken (for instance: another failure occurred or difficulties with the alternative train schedule)?
- Priority level 5.
    - Is the number of hindered passengers per hour as expected?
    - Other unexpected negative side-effect of the actions taken (for instance: another failure occurred during the day or difficulties with the alternative train schedule)?

Based on the answers to these questions it can be determined whether or not the goal was reached and whether or not actions have to be taken in the next phase.

Act: This step is necessary for the improvement of the process over the long term. For this situation, the evaluation done in the last phase, the check phase, might result in some necessary actions to be taken. This is especially true when one of variables is not as expected or if there are any negative side-effects following from the questions asked in the last step. If one of these things is the case, the following steps have to be taken:

1. Determine why the value of the variable is different as expected.
2. Determine whether or not this could have been prevented.
3. Look at the planning phase and see if, when you would have known the different value, you would have gone for another decision.
4. Take actions to be prepared for a deviation the next time such a situation occurs.

An overview of the steps is shown in Figure J.2.

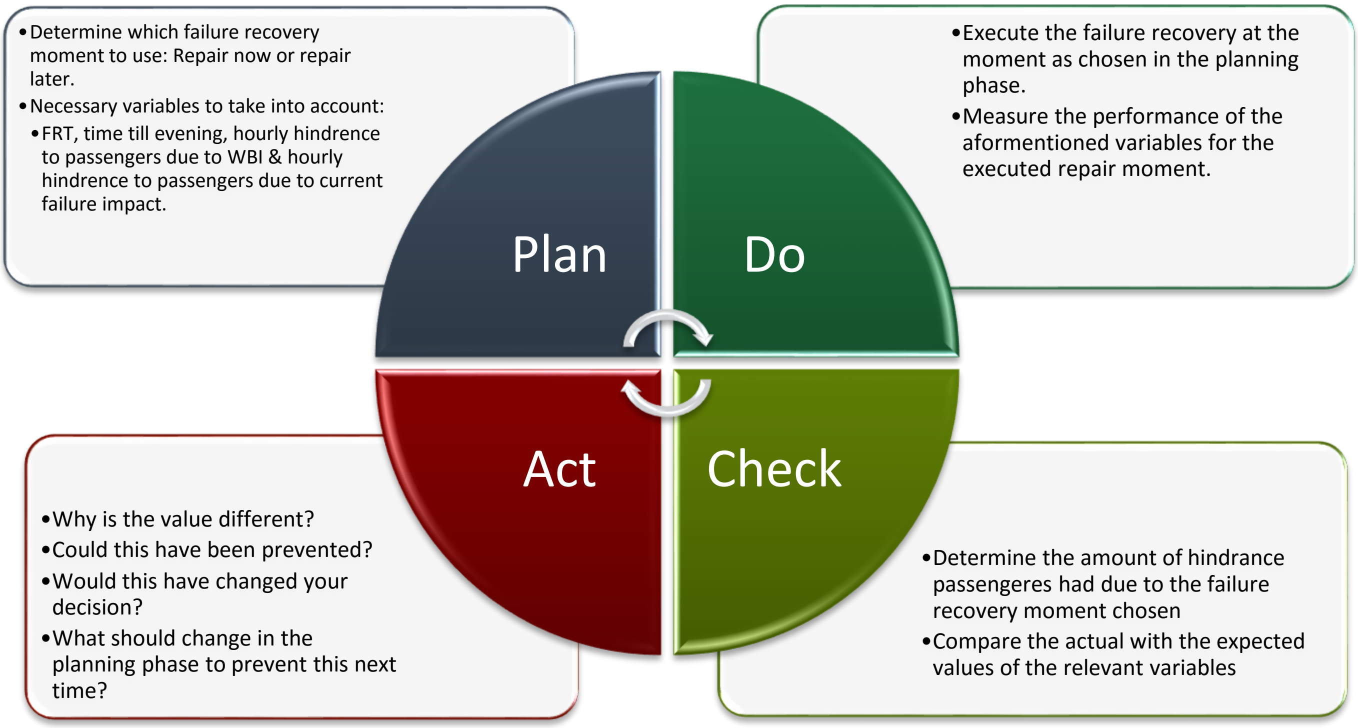


Figure J.2: The Deming Circle for the FRM-Decision-Making Process.

J.4. Conclusion

It is important to choose the right priority level and thus: the right failure recovery moment. In general, the decision is between either a repair during the night or a repair before the night at a time deemed suitable by the traffic control. An incorrect decision results in multiple trains being delayed while this is not necessary. Because of this, an unnecessary amount of passenger are hindered.

There are two combinations of variables influencing the decision between a repair in or not in the night. First: the FRT and the time until the evening and second: the number of passengers influenced by the WBI and the number of passengers influenced by the failure. The first variable of each of the pairs is part of the decision to repair the failure during the day while the other variables accompany a failure during the night.

With the use of the Deming Circle and these variables it is possible for the traffic control to evaluate their decisions and improve the decision making process. This enables them to make the decision with the least amount of impact of the passengers.

A representation of the execution of every step of the Deming Circle is shown in Figure J.2. As shown in the figure, during the circle the expected values of the variables named above should be compared with the actual values. These comparisons should be used to identify the reasons for the deviation, the effects this has on the decision taken and how this influences the future decision-making when the same situation occurs.