SMART MAINTENANCE AND THE RAIL TRAVELLER EXPERIENCE

Deliverable D2.3: CBM-Model Case Study Reports

Due date of deliverable: 30/09/2019

Actual submission date: 22/10/2019

Leader/Responsible of this Deliverable: António Andrade (IST)

Reviewed: Adam Bevan (UoH) and Daniel Johnson (ULeeds)

Document status		
Revision	Date	Description
D1	22/07/2019	Draft structure for partner contributions.
D2	31/07/2019	Preliminary contribution from IST.
D3	20/09/2019	Consolidating contribution from IST.
D4	25/09/2019	Contribution from HUD.
D5	27/09/2019	Review and final contribution from IST.
D6	17/10/2019	Last review.

Project funded from the European Union's Horizon 2020 research and innovation programme		
Dissemination Level		
PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

REPORT CONTRIBUTORS

Name	Company	Details of Contribution
António Ramos Andrade	Instituto Superior Técnico, Universidade de Lisboa	Chapters 1, 3 and 4.
Adam Bevan, Xiaocheng Ge and Farouk Balouchi	University of Huddersfield	Chapters 1, 2 and 4.

EXECUTIVE SUMMARY

The report presents case studies on the application of Condition-Based Maintenance (CBM) techniques for the prediction and optimisation of rolling stock maintenance activities.

Two main groups of case studies are explored: i) case study A: various vehicle systems and running gear and ii) case study B: wheelset maintenance. In case study A, a wide range of vehicle systems and components are analysed, using diagnostic data from Shift2Rail members and partners in IMPACT-2 project, to demonstrate the use of prognosis techniques in a CBM model. In case study B, the railway wheelset is the main component under analysis, where condition data from Fertagus train operating company has been used to demonstrate the application of statistical and survival modelling and a Markov decisions process to support the maintenance planning and optimisation of wheelsets. .

The report is structured as following: an introduction on the CBM model is provided in chapter 1, briefly reviewing techniques to support smart maintenance discussed in previous deliverables 2.1 and 2.2; in section 2 the vehicle systems and running gear case studies are explored using different statistical and machine learning techniques to analyse diagnostic data from rolling stock systems, providing a basis to support prognosis of the condition of several subsystems in vehicle systems; in section 3, the Fertagus train operating case study is presented, focusing on the wheelset maintenance component, using Linear Mixed Models, Survival Models and a Markov Decision Process (MDP) approach to derive an optimal maintenance decision map. This decision map is used to trigger maintenance decisions. Moreover, decision support models, i.e. a tactical maintenance planning model and an operational maintenance scheduling model are used to assess maintenance feasibility of a maintenance strategy. Both models are adapted to the Fertagus case study for a set of typical maintenance activities, including wheelset turning activities. A discussion is also provided on the precision associated with laser inspection, compared with manual and turning equipment. In section 4, a discussion on the CBM implementation is provided, pointing out the main lessons learnt from the case studies explored in sections 2 and 3. Guidelines and barriers to implementation are also identified and a final subsection on highlights of data inputs for impact assessment is also given. Finally, the main conclusions of the document are pointed out, showing the potential applicability of the CBM approach in other railway systems.

The application of the CBM techniques and architecture to a wide range of vehicle systems and components using real data from specific train fleets showed that such approaches and techniques can support maintenance decision as well as prognosis and failure prediction. It also allowed to discuss potential impacts and benefits using life-cycle techniques towards more informed condition-based maintenance regimes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

As discussed in deliverable D2.2, comprehensive maintenance decision-making processes have been developed in the transport sector and the MSG-3 methodology is a well developed process in the aviation industry. Such concepts and methodologies had been explored and discussed in deliverable D2.1, with an example case study on a typical system of an airplane. Deliverable D2.2 concluded that the MSG-3 methodology could potentially provide a useful basis for the definition and development of maintenance actions to support a condition-based maintenance (CBM) strategy for rolling stock.

The implementation requires applications of techniques from data science and condition monitoring to support the CBM system. In the system, two main goals are modelled: i) condition monitoring (CM), which mainly consists of data acquisition and storage of variables to describe the actual condition of a component/system and allow the identification of the root causes of system failures and ii) maintenance decision supporting, which consists of failure prognosis, development of guidance and evidence to support maintenance decisions, and continuous improvement of maintenance decisions and their impact from other perspectives, e.g., prolonged asset life and savings in life-cycle costs.

A comprehensive procedure for condition-based maintenance of rolling stock was described in deliverable D2.2, describing the proposed techniques to support predictive and preventive maintenance and the major tasks in the CBM system, namely:

- Condition monitoring and data collection;
- Data cleansing, pre-processing and signal processing;
- Feature selection and extraction;
- Statistical modelling;
- Fault diagnosis and prognosis;
- Maintenance decision, planning and optimisation.

Data processing and feature extraction in complex systems such as railway vehicles requires using different statistical models to better describe trends and predict the behaviour of subsystems or components, decomposing many times such information in frequency domain. Several techniques were described in deliverable D2.2, with discussion on data visualisation and data driven prognostic methods (e.g. Markov Chain models, Hidden Markov models).

In parallel, three critical challenges for CBM to be successful were identified. First, a system health indicator must be determined or defined, using for example dimension reduction techniques (e.g. Principal Component Analysis). Secondly, accuracy and interval of condition monitoring should be ensured, requiring inspection and sensing technologies (e.g. laser equipment) and using predictive techniques such as Artificial Neural Networks. Finally, condition limits must be defined, which might change during the life-cycle or operation time of the component or subsystem.

Supporting maintenance decisions, as the second goal of a CBM system, requires analysis techniques that can derive optimal maintenance strategies given the expected evolution/transitions of the condition of a given subsystem or components. In that sense, Markov Decision Processes are controlled stochastic processes that can be applied to derive an optimal maintenance action map. This is achieved using wheelset condition data for the case study of Fertagus train operating company in Section 3.3 of this report. A tactical maintenance planning model and an operational scheduling model are also applied to that case study, supporting the assignment of maintenance activities to each train unit, given several constraints (e.g. depot).

The case studies presented in this final WP2 deliverable demonstrate the application of the developed CBM techniques for the prediction and optimisation of maintenance for a range of rolling stock components/systems.

2. USE OF ON-TRAIN DIAGNOSTIC DATA IN CBM

In rail operation, maintenance is a significant proportion of the total operational cost. In the SMaRTE research project, we have explored several approaches of CBM using prognostic models to support maintenance decisions based on the predicted condition of a component/system. Although an ideal CBM approach is directly based on condition data (e.g. data which describes the actual condition of the system/component at a particular period in time), and in which a prognostic-based decision support for CBM is not an extensively explored area, the objective of the research described in this section of the report is to develop methods to deal with the specific challenges associated with the use of diagnostic data for CBM, such as variations in data quality. This will be used to predict the occurrence of specific system diagnostic events (including warning and failures) and to continuously update maintenance-related decisions.

Overall, a framework of a CBM process (shown in Figure 1 below) has been developed which focuses on dealing with real-time diagnostic data which is gathered in high frequency, developing prognostic models for the estimation of the remaining useful life (RUL) or remaining life distribution (RLD) and providing recommendations for future CBM. In deliverable D2.2, SMaRTE already reviewed some techniques (including some commonly used statistical analyses, machine learning techniques and Markov-based optimisation of maintenance plans). The focus of this deliverable is how these techniques can be applied into real systems. As highlighted in the Figure 1, the case studies will demonstrate the processes in two stages of a typical CBM process, namely Prognosis and Decision support.

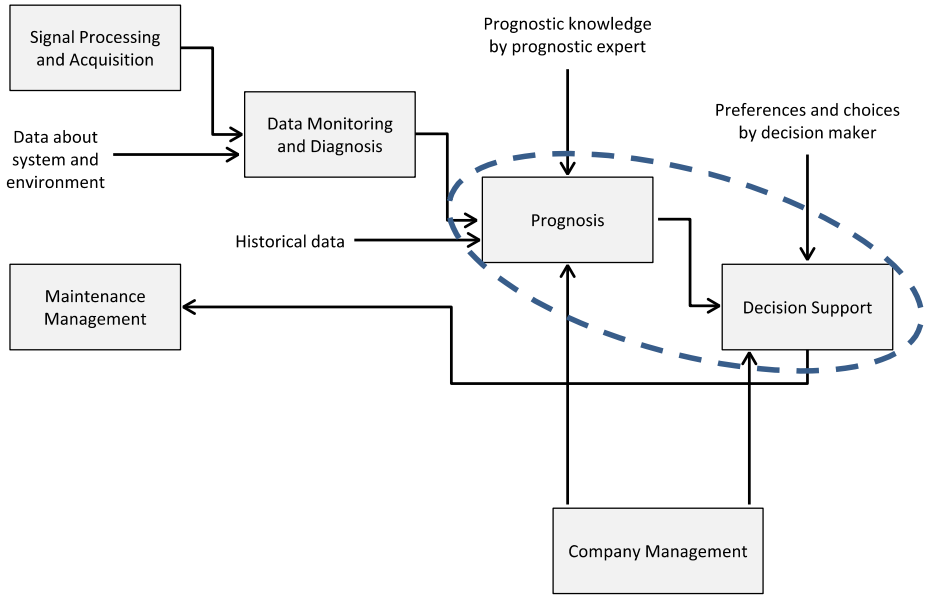


Figure 1: Proposed CBM process (adapted from Voisin et al. 2010)

The implementation of the prognostic-based approach in different CBM case studies is not straightforward because of the diversified characteristics of data in each system. Moreover, our method often needs to be combined in order to cover the prognostic and decision-making requirements of CBM applications. Hence, there is a need for systematically describing the methods' characteristics and supporting the selection of the most appropriate combinations for the needs of specific CBM applications.

Hence, research works dealing with decision support for CBM usually develop a model of analysing and processing the historical data available, apply this prognostic model online on the real-time data streams and, based on the derived on-line prognosis, develop decision methods in order to provide prognostic-based recommendations. Specifically, prognostic methods are applied first and decision methods are then used in order to provide maintenance recommendations as shown in Figure 1.

In this first case study; diagnostic data from a range of vehicle systems (e.g. automatic sliding doors, air conditioning unit, braking and traction systems) have been analysed to demonstrate the use of prognosis techniques within a rolling stock maintenance environment. These techniques can be combined with the condition-data driven techniques and decision support processes described in Part II of this report to developed a CBM-model for rolling stock.

2.1 ON-TRAIN DIAGNOSTIC DATA

Unlike fault diagnosis, prognosis is a relatively new area because traditionally prognostics was viewed as an add-on capability to diagnosis, however it has become an important part of the CBM system. Prognosis assesses the current and historical status of a system and predicts its remaining life based on features that capture the gradual degradation in the operation of the system. Predictive capability is critical to improve availability, plan successful missions, schedule maintenance, and reduce maintenance costs.

Ideally, when designing a condition-based maintenance system to provide early warning of failures and inform maintenance planning, it is clearly advantageous, and a better use of resources, to have the vehicle diagnostic system obtain information about the failure symptoms by using sensors already installed on a vehicle to capture the failure information. Retro-fitting sensors to existing vehicles can be prohibitively expensive requiring integration into a system that, for all intent and purposes, are likely closed for such upgrades. At this stage it would be beneficial, and which is presently not common, to add higher layers of prognostics to on-train systems to take full advantage of the rich data structure and inherent engineer domain knowledge that is already present and accessible via the on-train diagnostic system.

Modern rolling stock are typically equipped with an on-board diagnostic system that continuously monitors the operation of critical systems, as illustrated in Figure 2, and generates specific event data when abnormal operation or a specific event is observed. The system has what can be described as an event-driven architecture. This is programmed to record fault diagnostics information efficiently. Typically, diagnostic information is used retrospectively when the vehicle is

called for maintenance actions; however, if the data can be accessed in real-time it could, feasibly, be used to predict an impending failure with sufficient response time to allow remedial action to be planned before the event becomes terminal.

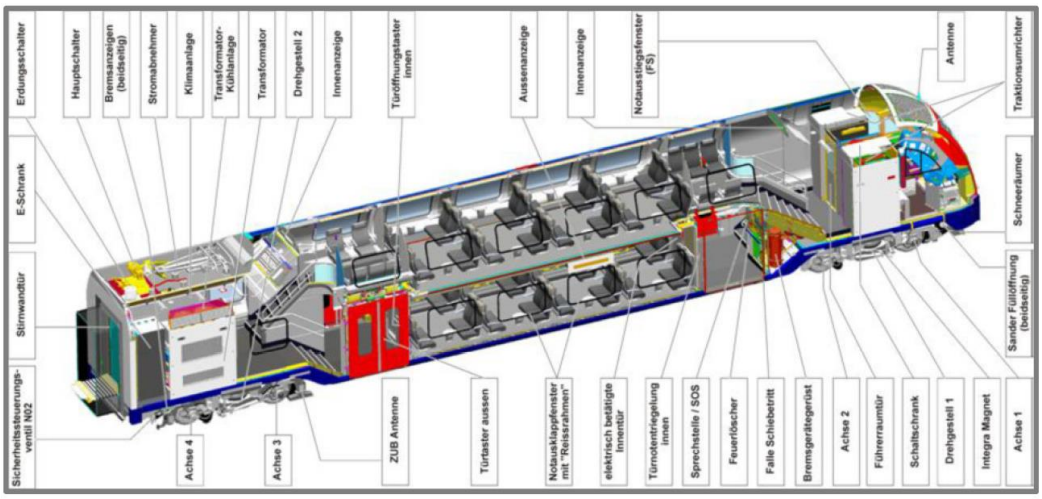


Figure 2: On-train diagnostic system (Source: Schulte-Werning et al. (2016)).

Off-train monitoring configurations have significantly different hardware and programming characteristics to those of on-train systems. Data capture and processing is usually implemented in a manner very specific to the component being monitored. Sensors, microcontrollers, sampling frequencies, duty cycles, memory and data extraction are all designed to monitor the condition being observed. Power, in these circumstances, is usually limited enforcing stringent data transmission and power supply constraints.

2.1.1 DATA COLLECTION AND FORMATTING

On-train monitoring systems, as previously mentioned, are event-based with differing operational constraints. Power is not usually a limited commodity but other parts of the monitoring system, such as memory and sensory data, can be limited due to the wide range of vehicle systems being observed. The on-train diagnostic system is capable of capture over 5000 diagnostic codes that, in combination, are simultaneously gathering operational data on a wide range of the vehicle systems from the traction and braking system to the air conditioning unit and automatic doors.

Modern train are equipped with a diagnosis system, which consists of a range of logical devices and sensors/actuators in order to collect and transmit current faults/events and environment variables to a data logger via the multi-function vehicle bus (MVB). The systems of interest are monitored by the diagnosis system throughout a historical period of observation. Among the data captured and stored in a diagnosis system, it is ideal that both sensor and event data are available for prognosis. However, it is more often that only event data is available. The severity of an event and/or how safety critical the system is will dictate how the system records the data. These event data herein are defined as a collection of data items containing at least a time stamp, a failure/event code and eventually a descriptive text.

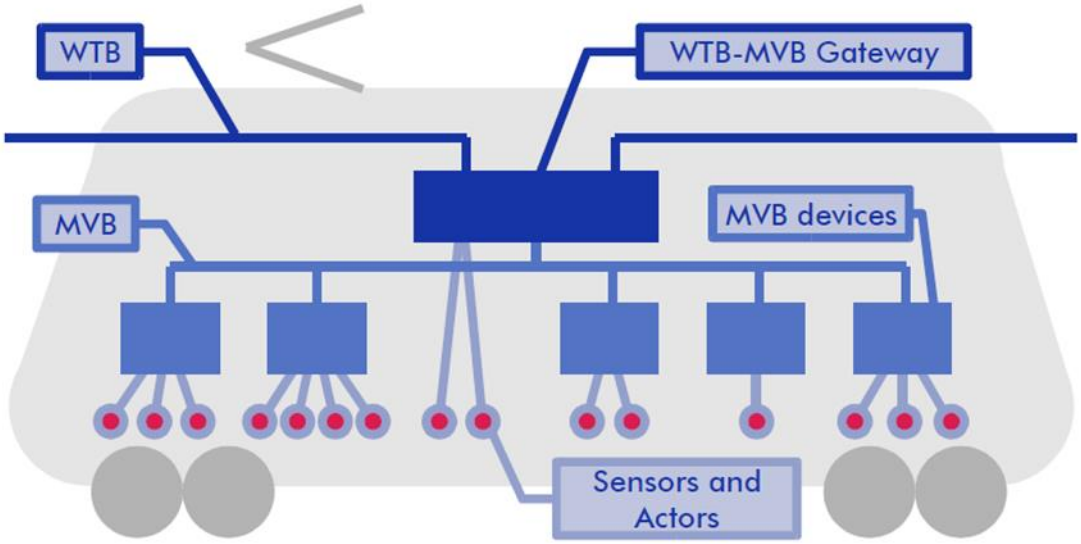


Figure 3: Overview data generating resources.

It was identified early in the SMaRTE project that there were going to be some difficulties in obtaining condition and diagnostic data from fleets operated by project partner LUL. To mitigate this risk, data sources provided by IMPACT-2 WP6 partners were utilised in the research (under Shift2Rail Cross Cutting Activity Grant Agreement 777513).

Data from a range of vehicle fleets were provided by IMPACT-2 partners to support the development of analysis techniques for the prediction of future component/system failure. This mainly included diagnostic data, however some sensor data and information on maintenance activities were also provided. The data provided by IMPACT-2 covered regional, suburban, double-decker and high-speed electric multiple units (EMU) and a number of systems (e.g. traction/braking system, doors and air conditioning unit). Following some initial problems related to the collection, formatting and quality of the data, the data was pre-processed using statistical analysis techniques and cross-referenced with maintenance records as detailed in Section 2.2.

2.2 ASSESSMENT METHODOLOGY

The overall process of prognostic-based approach adopted in the research is shown in Figure 4 below. This methodology has been applied to the assessment of diagnostic data for a regional and high-speed EMU in Section 2.3 and 2.4.

To support this a range of data processing and analysis techniques were applied to the diagnostic data.

I. Selection of failure(s) of interest and relevant diagnostic codes based on:
    a. experience of depot staff and/or known codes which result in maintenance interventions

b. design of on-board diagnostic system architecture (e.g. failure of safety critical components)
c. statistical analysis of diagnostic codes / events (step II below)

II. Statistical analysis of diagnostic codes to characterise failure(s) of interest
a. frequency of occurrence, time activated, cumulative distribution
b. relationships between diagnostics codes and measured environmental variables

III. Prognostic model to forecast future activations
a. linear/non-linear regression
b. neural networks

IV. Early warning of potential failure
- support maintenance planning (rather than corrective/unscheduled maintenance)
- maintenance management system

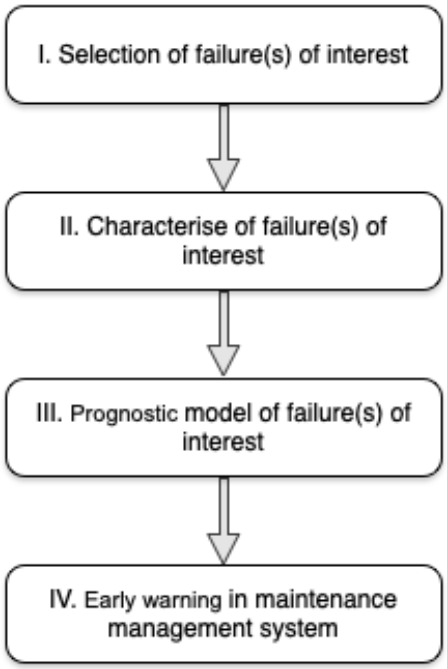


Figure 4: Prognosis-based assessment methodology

2.3 CASE STUDY OF REGIONAL EMU

The first type of train selected in this case study is a regional EMU, known as Class 440 Coradia. These were manufactured by Alstom and operated by Deutsche Bahn (DB) since going into service over a decade ago. The Class 440 traction control unit (TCU) and retractable door step were initially selected following discussions with DB Regio AG's product line support and maintenance departments, however the focus of this case study is the TCU.

Work instructions for scheduled preventive maintenance exist for both components in the current maintenance plan of DB Regio AG. These work instructions contain defined work contents, which must be carried out at fixed mileage limits within planned maintenance periods. The work instructions with defined work contents, limit values to be complied with and mileage limits have

thus far been based on manufacturer specifications, calculations, experience from maintenance and operation etc. The work instructions apply to the entire fleet of the relevant vehicles of an RU or groups of vehicles operating under the same conditions. The work instructions, including all specifications, therefore do not take into account the individual condition of a specific vehicle component.

The planned introduction of a data-based monitoring system is an attempt to determine the actual condition of individual components and thus define maintenance requirements individually for the components. The data and signals already available in the vehicle or component are used for this purpose. They are assigned to the respective functions or possible causes of faults in order to be able to make statements about the actual condition of the component using analysis methods and pattern recognition. The ultimate goal is to establish condition-based rules for the work to be carried out and thus to partially or completely transfer the original work instruction from the above-mentioned scheduled preventive form into a condition-based maintenance work instruction.

2.3.1 SELECTION OF FAILURE/EVENT OF INTEREST

A rail vehicle consists of many systems which are monitored for the purpose of operation and maintenance. The selection of appropriate systems was performed in consideration of the following criteria:

- Relevance for maintenance (scheduled or unscheduled/corrective)
- Relevance for safety, reliability or comfort of the passenger trains
- Availability of relevant data for CBM application

Considering the recommendations from DB and initial interrogation of the entire set of event data, the TCU was selected as it is subject to frequent unplanned/corrective maintenance and a certain amount of diagnostic and maintenance data was available for the case study.

In the data set, the following field data available for the prognosis:

- Event logger (fault records reported during operation and resulting in maintenance):
  - Reference period of 4 years: 2015 – 2018
  - Number of trains: 152
  - 32 types of reported events for TCU
  - Total number of reported events: 2401 for TCU
- Maintenance records (completed maintenance activities):
  - Reference period of 4 years: 2015 – 2018
  - Number of trains: 152
  - 59 types of corrective maintenance actions for TCU
  - Total number of corrective records: 938 for TCU

There will be many events occurred during the operation of a system. In this case study, there are 32 types of events for the TCU. The objectives of “Selection of failure(s) of interest” are: a) to have an overall understanding of the system in terms of the occurrence of failures/events; b) to select candidates of further analysis and prognosis. There are many approaches in order to select the

failure(s) of interest. Recommendations and suggestions from maintenance engineers is the most important piece of information for our analysis. DB's maintenance engineers were consulted to identify which failure(s) are most concerning (from a safety and maintenance resource perspective). In addition, maintenance/failure reports from the depot were also critical for us to have a focus since there may be hundreds of faults on a fleet during an inspection cycle. The maintenance/fault report from DB has a following structure:

Field Name	Description	Example
Erledigungsstatus aus ZIIS_AV_P_F_K	Completion status	E
Status des Schadens	Degree of damage	F
Erfassungsdatum	Entry date	04/01/2015
Erledigungsdatum	Completion date	06/01/2015
Schadcode	Fault code	FA38
Fehlerklasse	Fault class	3
Materialkennzeichen	Material ID	J
Schadcodetext	Fault text	ASG 1 gestört
Kurztext	Short description of fault	Antriebsteuergerät, Störungssuche
Langtext	Long description of fault	Speedsensor Radsatz 4 defekt (=WR1 Speedsensor2)
IH-Werk	Depot ID	12LR
Aktuelle Fahrzeugnummer	Vehicle ID	94800440203-8
Arbeitsvorrat-Nummer	Maintenance activity number	225897054
Codegruppe-Arbeit	Activity code	TBGCBCT

By analysing this fault reports, the most significant faults can be identified which result in unscheduled/corrective maintenance prior to the regular service interval (e.g. every ~52km).

However, one of the most effective approaches for the selection of failure(s) of interest is to analyse the historical data. Figure 5 shows the occurrence of all failures/events over the period of the observations. This approach can provide useful information to:

- Identify the coverage of the data over the analysis period. As discussed in previous deliverables, data availability is one of project risks and limitations of CBM. This initial analysis helped us to understand the coverage of historical data and the selection of data for the further analysis and prognosis.
- The potential indication of a group of failures/events which are highly correlated. Based on experience, events/failures of a system may have some interdependencies in general.

In Figure 5, it can be seen that a) there are some gaps in the dataset, for example there is data missing for the first and third quarter of 2016, however we have some good continuous data over 2015. For our analysis, it is better to work on data from 2015 for the prognostic model; b) The occurrence of failure codes 19048, 19049, 19051, 19052, 19056, 19057 shows some potential correlation in the chart. This requires some further analysis to understand if these failure codes are proposed to be correlated (by the definition and design of diagnosis system) or there is a "true" dependency among these codes.

2.3.2 CHARACTERISATION OF FAILURE/EVENT OF INTEREST

After consulting the maintenance engineers from DB, a group of failure codes (i.e. 19048, 19049, 19051, 19052, 19056 and 19057) were selected which represent an event with the inverter “inverter disturbed” accompanied by the “disturbance of ASG traction control unit”. As designed in the diagnosis system, it makes sense that these codes are strongly correlated which means one event occurs and the other events will be captured in a short period. However, further analysis of this correlation is needed, i.e. which event is primary and which are just followers.

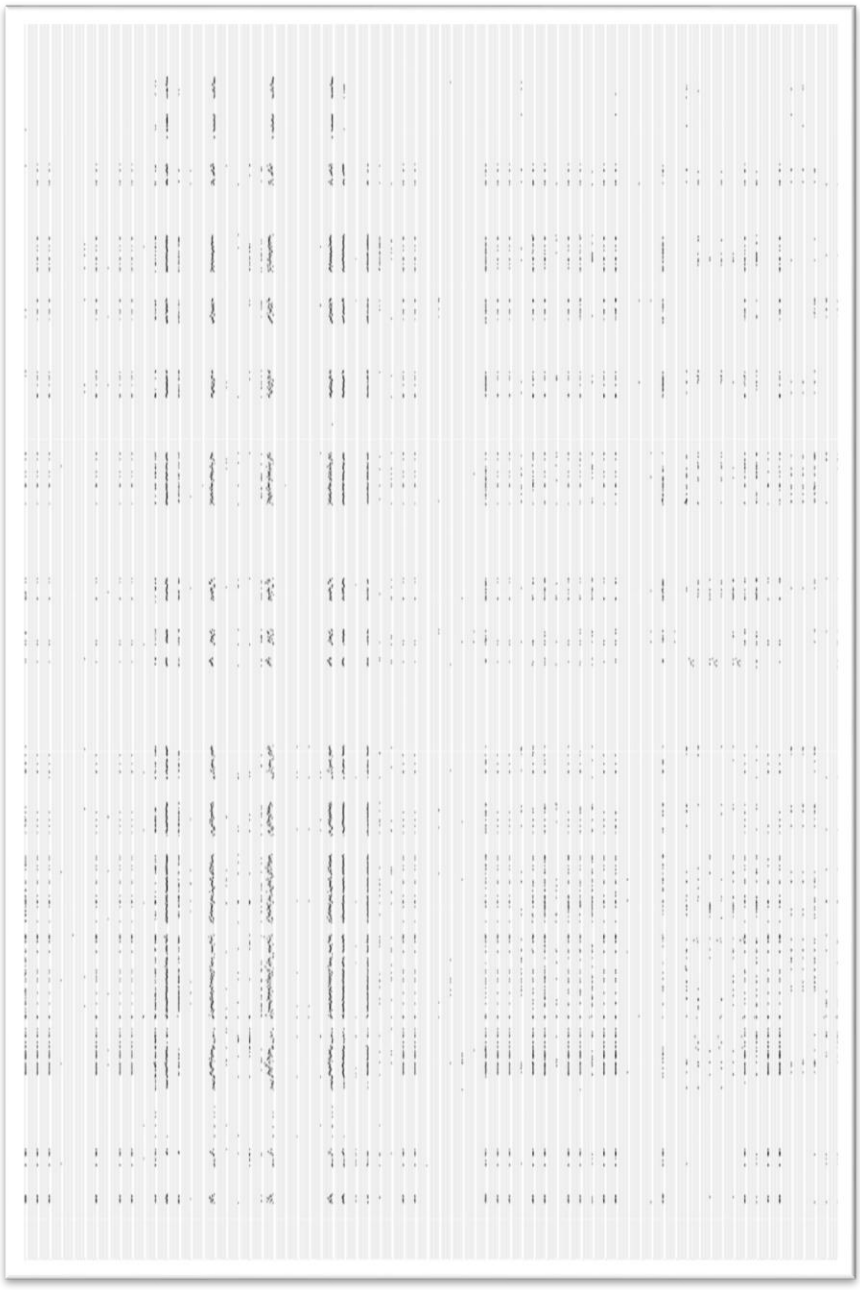


Figure 5: Failure codes occurrence

Meanwhile, we also obtained information from maintenance engineers that failure code 19056 (“wheel diameter difference too high”) presents the reason for a traction disturbance, which would be a clear indicator for maintenance. The occurrence of specific codes is coloured in red to highlight 19056, which is the most significant failure in the group. At step II of Figure 4, “characterisation of failure(s) of interest”, there are two tasks to do to identify the relevant characteristics of the occurrence of these codes:

- The first task is to visualise the timeline of occurrence of codes over the period, shown in Figure 6. This can provide information of when the event is on and how long the event lasts.

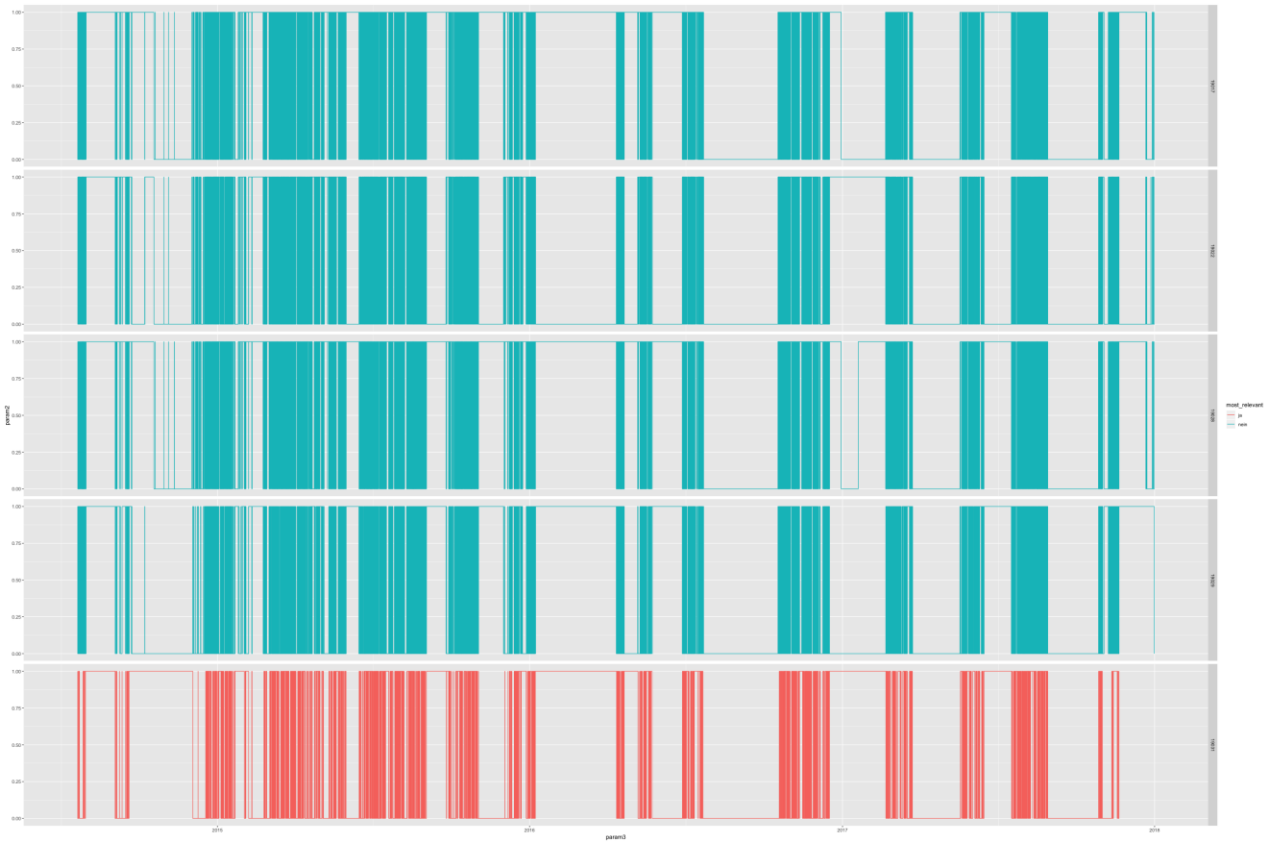


Figure 6: Visual timeline of code occurrence

In the case study we found that these codes/events are almost synchronized – they arise and disappear at the same time (or the delay is not captured). Ideally, we expected that such an analysis can answer the question of which event is the primary and which events are the followers.

- The second task is to visualise the trend of occurrence, shown in Figure 7. It can provide some information of how fast the events’ occurrences are increased over the analysis period.

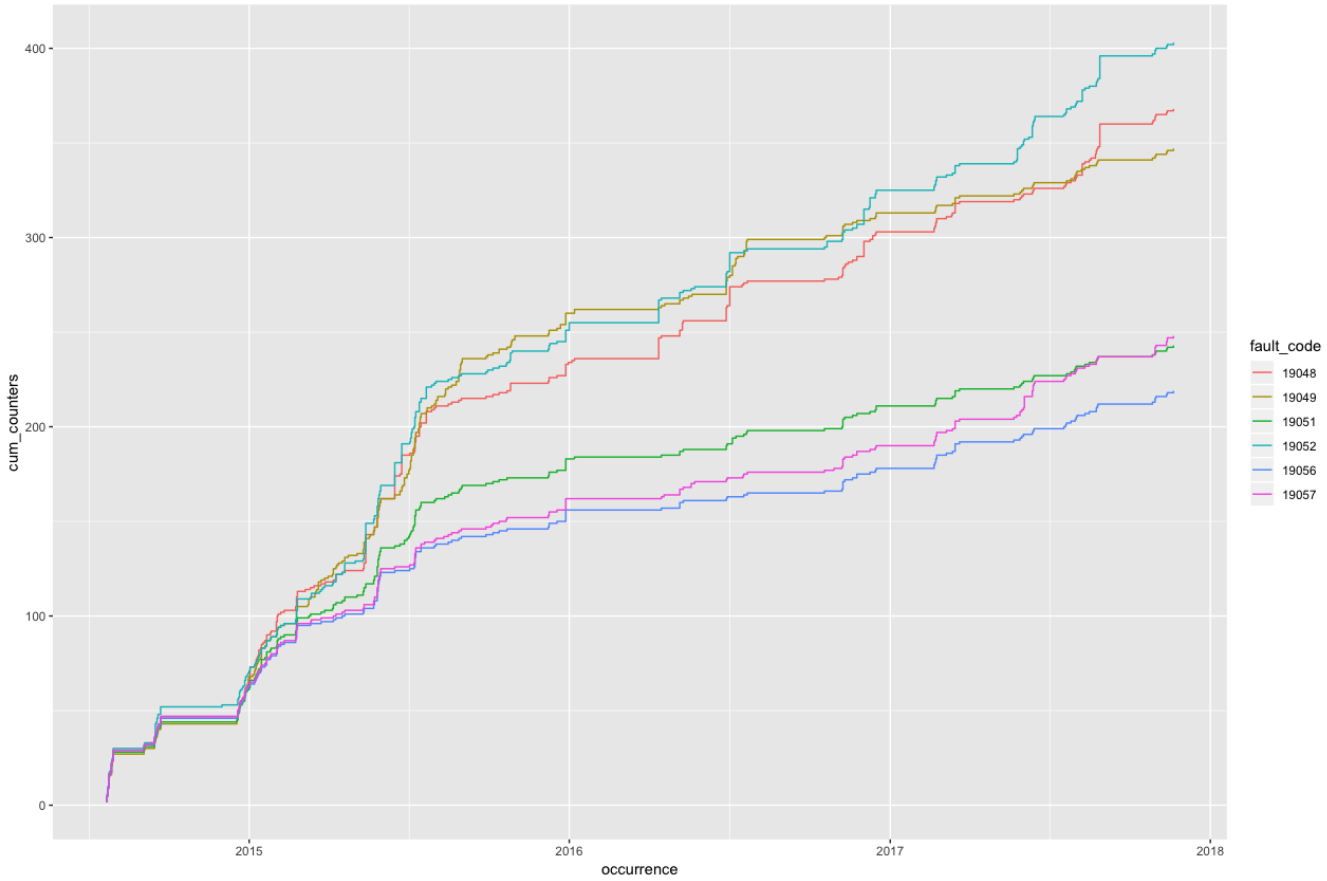


Figure 7: Trends in failure code occurrence

Most importantly, the projection here tells us that the occurrence within the group splits further into two groups. The upper part of the plot indicates the trend for inverter no. 1 (code 19048, 19049, and 19052) which is shown to be more prone to failure than inverter no. 2 (code 19051, 19056, 19057). It also provides us the information that the rate of occurrence increases after a certain point and are roughly at the same rate. In reliability engineering, we know that the system's failure rate is a bit higher than the rate in the later stage of useful life (the bath-curve). So it is reasonable to believe after the first period of time these inverters were working in a stable condition. Also, within the time window of the historical data we did not observe a dramatic change in occurrence rate.

2.3.3 PROGNOSTIC MODEL OF FAILURE/EVENT OF INTEREST

A reasonable prognostic model is built on the rigorous analysis of historical data. Using this prognostic model, we are then able to forecast the occurrence of an event/failure. In this case study, we have tried two types of predictions: the prediction of occurrence based on a linear regression model on accumulated occurrences and the prediction of occurrence based on a recurrent neural network model.

- Prediction of occurrence based on a linear regression model on accumulated occurrences

The fundamental assumption of the prediction using a linear regression model is that the occurrence of an event is distributed evenly over a period. In our case study, we identified that the total occurrences of events of interest increase steadily after the first period (i.e. after June 2015), and it is reasonable to assume that all these failures will arise constantly for the near future.

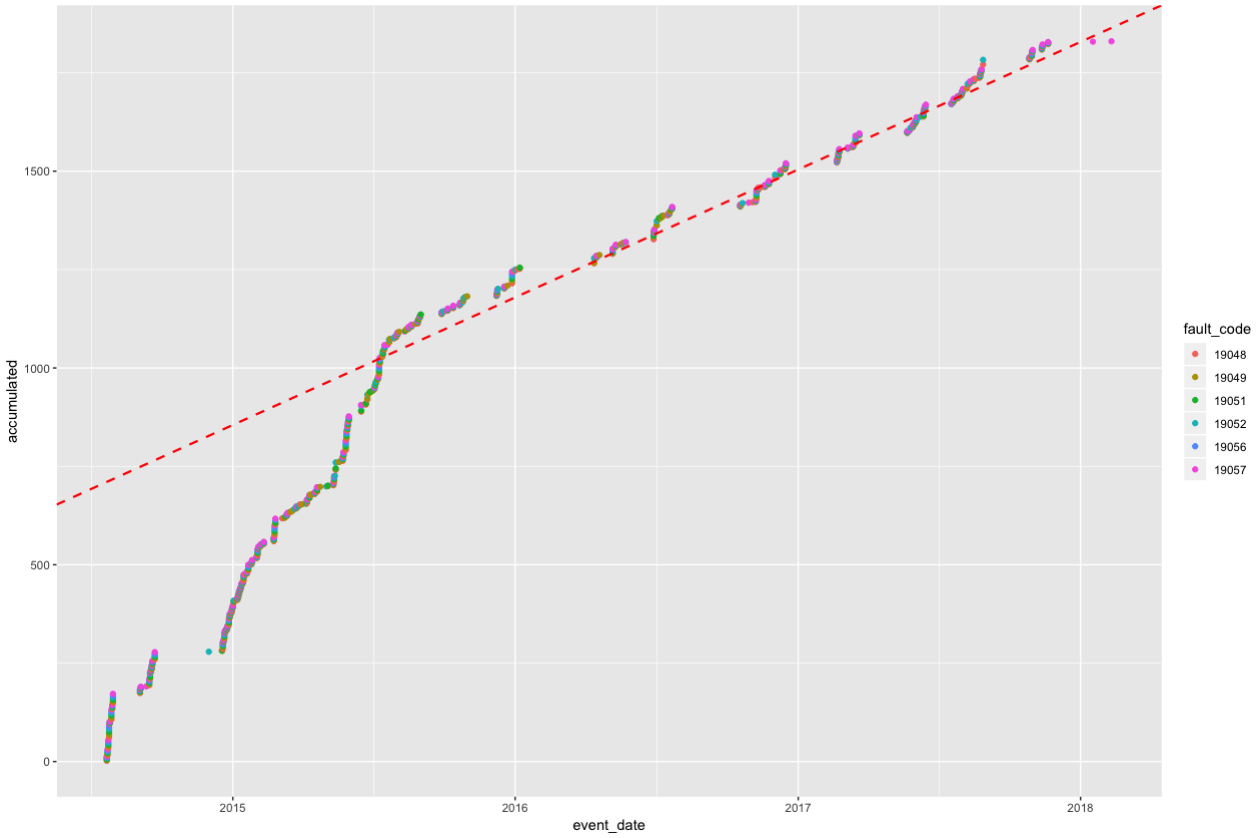


Figure 8: Linear-regression model

Based on the prognostic linear model, we can estimate when the total number of occurrences of these failures reaches a specific threshold, which determines when the vehicle needs to go to depot for service or maintenance.

- Prediction of occurrence based on a recurrent neural network model

Creating a recurrent neural network (RNN) prognostic model and using it to predict the occurrence of events is another approach that has been trialled during the research and included in this case study for both regional and high-speed EMU (Sections 2.3 and 2.4). Further details of the RNN model can be found in Section 2.4.5. Compared with prediction based on a linear model, the fundamental difference is that it is not based on the assumption that the events to predict will happen constantly over a period. Shown in Figure 9, the selected group of failures did arise irregularly over the period.

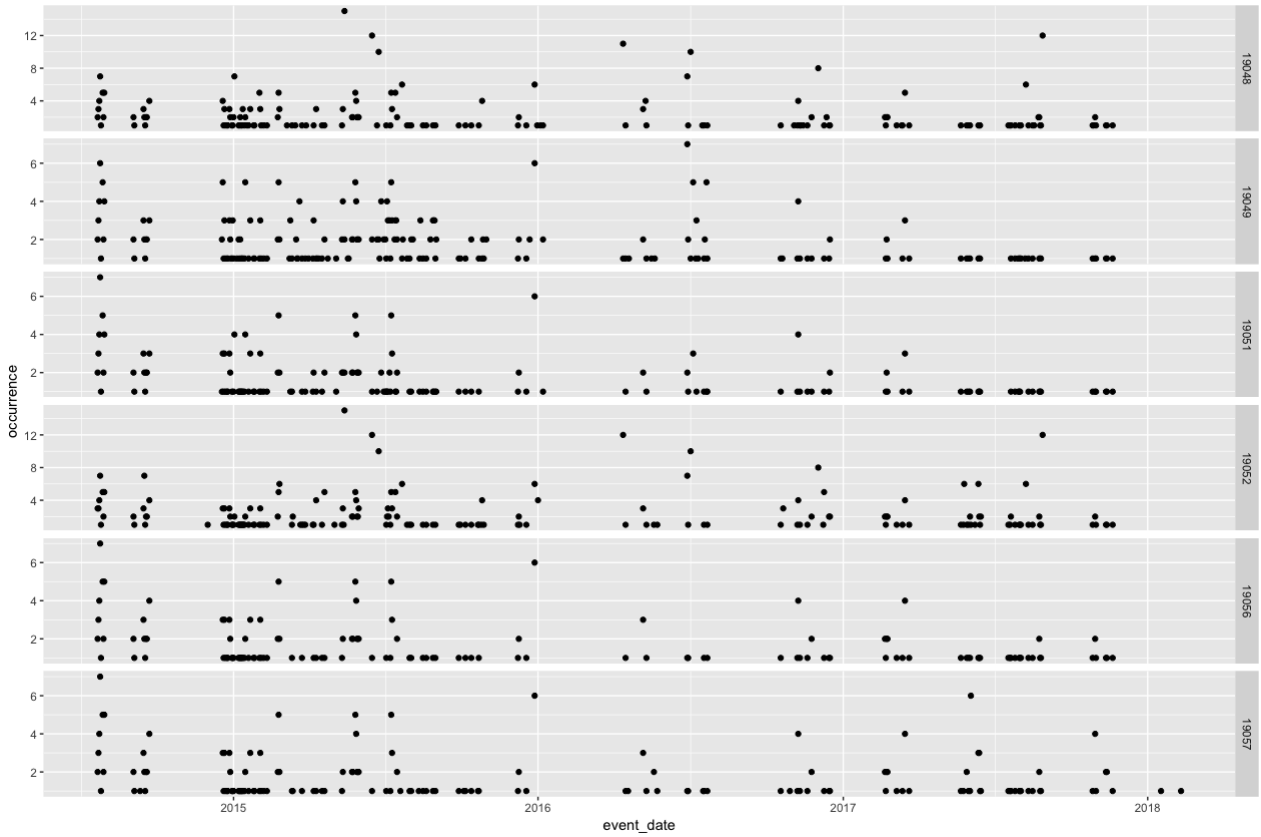


Figure 9: Irregular occurrence of failure codes

There are two approaches to predict the occurrence of failure codes: the first one is to predict the number of occurrences in next a few days; and the other approach is to predict when the codes will arise and disappear in next a few days.

1. Prediction based numbers of occurrence in next period

To predict how many times the codes will arise, we need to preliminarily process the data to know how many times the events are recorded in the historical data. The data is aggregated as summarised below:

Event_date	F_19048	F_19049	F_19051	F_19052	F_19056	F_19057
.....						
22/07/2014	2	2	2	3	2	2
23/07/2014	3	3	3	3	3	3
24/07/2014	4	4	4	4	4	4
25/07/2014	7	6	7	7	7	7
26/07/2014	1	1	1	1	1	1
28/07/2014	5	5	5	5	5	5
.....						

Meanwhile, we can calculate a correlation efficiency matrix to show how these failure codes are correlated in the history.

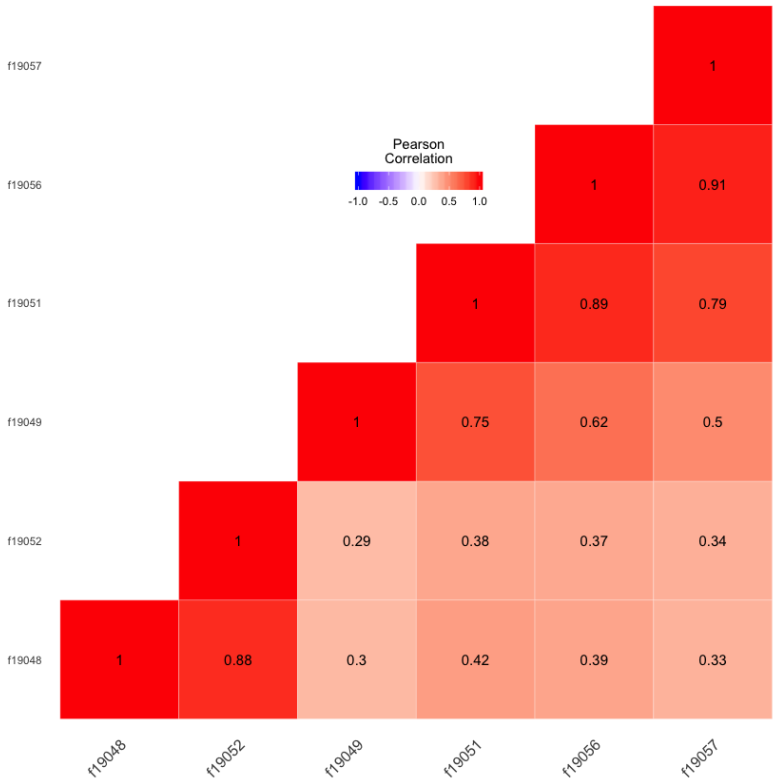


Figure 10: Correlation efficiency matrix

It has been identified that failure code 19056 is more correlated with codes 19049, 19051, and 19057, and slightly less correlated with codes 19048 and 19062.

After the statistical analysis, a RNN model was built to feed the historical data into the model so that the model can learn from the historical data and search for patterns. After cleaning the historical data (omitting the days that records are missing), we have 895 rows remaining.

Some intimal parameters for the neural network model are set as below:

- lookback = 28 — Observations will go back 28 days.
- steps = 1 — Observations will be sampled at one data point per day.
- delay = 5 — Targets will be 5 days in the future.

The historical data is scaled by using its standard deviation and mean.

Mean:

F_19048	F_19049	F_19051	F_19052	F_19056	F_19057
0.02369931	0.06957364	0.05434398	0.02230618	0.03565942	0.03378029

And Standard Deviation:

F_19048	F_19049	F_19051	F_19052	F_19056	F_19057
1.0313293	0.9966763	1.0266073	1.0215954	1.0307834	1.0312577

When using neural networks, we choose mean absolute error (MAE) as the loss function to train the neural networks. Then we can have a benchmark of prediction by calculating the MAE without any prediction. In the case of code F_19056, the MAE is 0.4757996. Because the historical data has been normalized to be centred on 0 with a standard deviation of 1, this MAE isn't immediately interpretable. It translates to an average absolute error of 0.4758 x F_19056's std: 0.5, which means the prediction is within an error range of ± 0.5 (the difference between the number of events recorded and predicted).

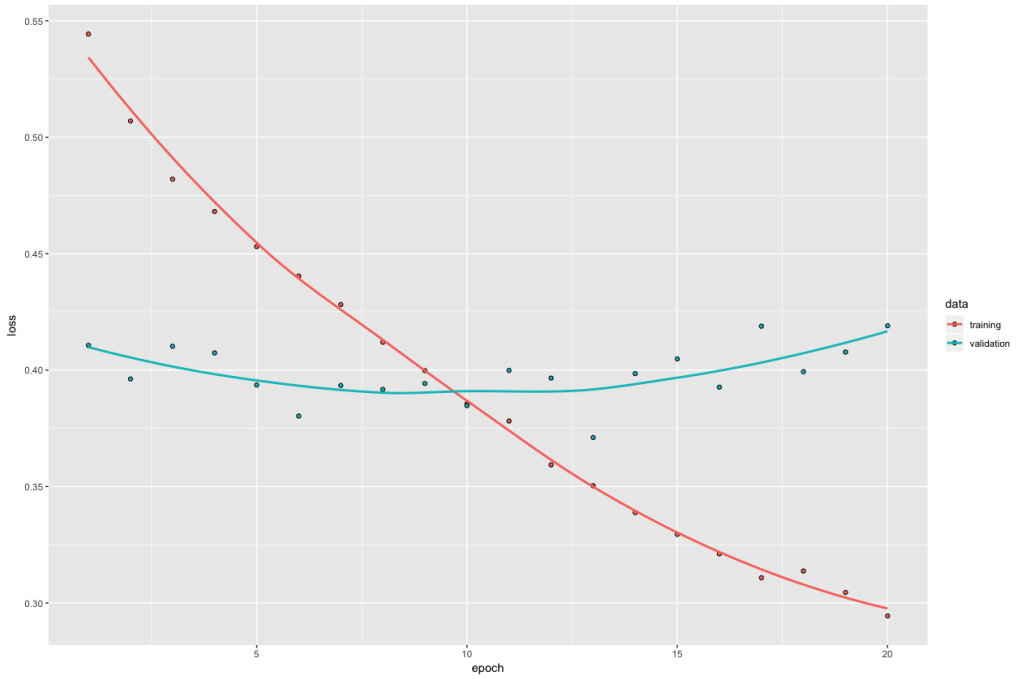


Figure 11: Validation of NN prediction model

Figure 11 shows the loss (MAE) of training and validation progress when the RNN is fitted to the data. In the dataset we feed into the RNN model, we split the first 70% worth of data for training, and use the remaining 30% of data to validate the model. In the Figure, we can see that the RNN model digested the training data quickly, and the MAE is dropped from 0.53 to 0.3 after 20 rounds; however, the validation based on the rest 30% data samples does not follow the trend with a loss of approx. 0.4.

After making the neural network model learn the historical data, the MAE of learning set is down to 0.2944, which is better than the approach without neural network model. The prediction results are shown in Figure 12 below.

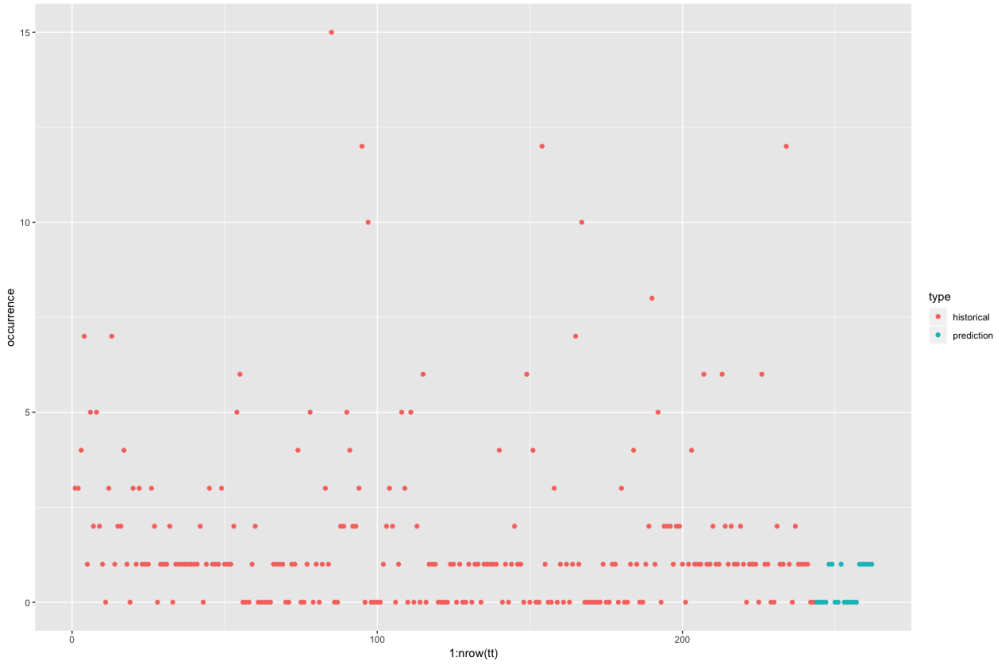


Figure 12: Prediction of failure occurrence

There are several interesting observations from the figure of RNN fitting history (Figure 11) and the prediction results (Figure 12).

- The loss in the training and validation data crosses at around 0.4 and the loss of validation does not follow the trend of the training data. This could be caused by several factors: 1) the RNN model is set to learn from 28 units of historic data in order to predict the situation for the next 5 units. However, the actual data may not have a periodical pattern; 2) the first 70% of the data may have different characteristics (e.g. different periodical pattern) to the remaining 30%. After further investigation, we are more confident that the results of our predictions are not as realistic as possible (although the actual MEA, e.g. the difference between the actual recorded events and the predicted number of occurrences, is nearly 1 i.e. if predicted event X will happen n times at a particular day and actually the number of event X recorded is between n+1 and n-1) because the data we are holding is not continuous so that the characteristics of the time series of event X will be different and therefore the RNN model cannot find the “real” pattern. This highlights the importance of data quality, especially the continuality of data, in event-based predictions.
- By scanning the entire data, a preliminary step is to manually explore any obvious periodical pattern in the data. So that the parameters (28 days and 5 days in our case study) can be selected wisely. One of reasons why our experiment in this case study does not provide a “good” enough result is that the selection of the parameters for the RNN model are not appropriate. It can be improved by changing the values of these parameters. However, several values of the parameters (e.g., 12 days and 3 days) were trialled, however we were still unable to show a good prediction due to the reasons identified above.

2. Prediction of when/if the codes will remain active of disappear in next a few days

As shown in the historic data, an event code can remain active or re-appear if the fault persists or disappear if the fault is corrected during operation. Based on the historical data, it was attempted to predict the occurrence of events based on when the events appeared and disappeared, as shown in Figure 13 below.

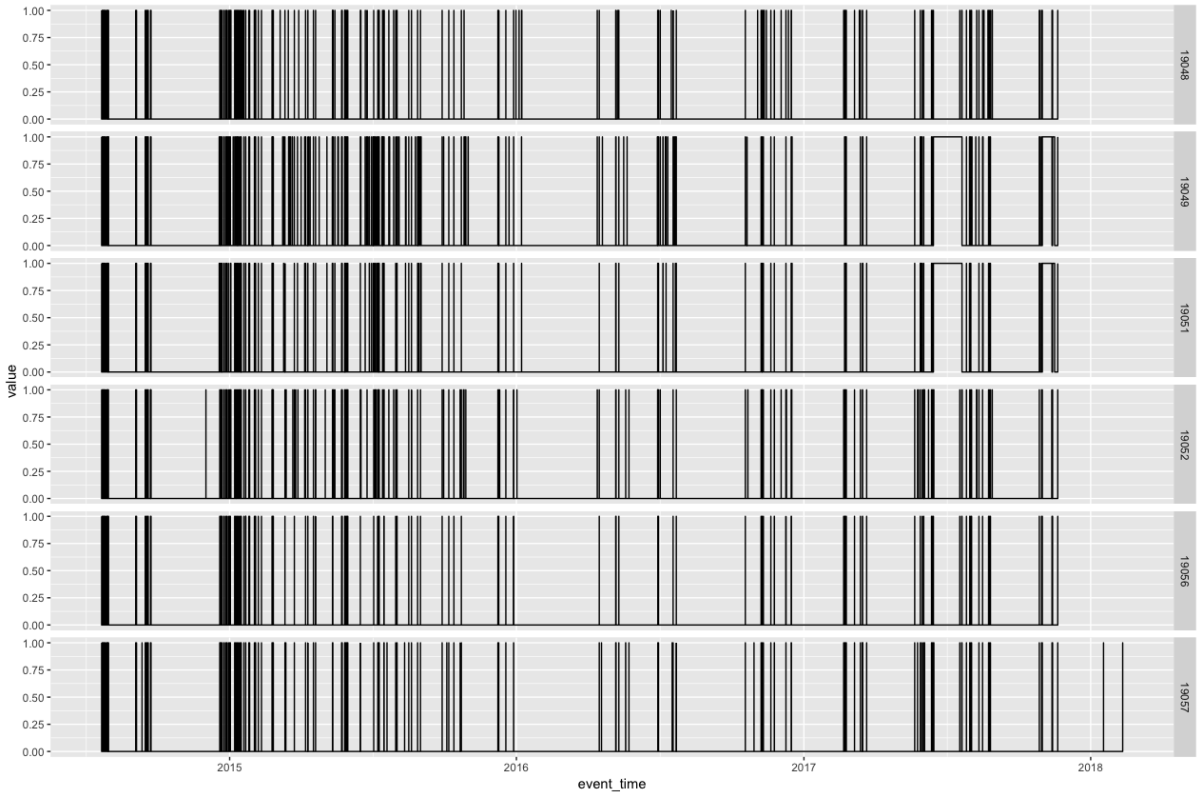


Figure 13: Occurrence of events from historic data

Because there are gaps in the data set, we attempted to clean the data in a different way, and the processed data looks like:

Event Date	F_19048	F_19049	F_19051	F_19052	F_19056	F_19057
2014-07-22T01:07:45Z	1	1	1	1	1	1
2014-07-22T01:08:29Z	0	0	0	0	0	0
2014-07-22T06:43:20Z	0	0	0	1	0	0
2014-07-22T06:49:11Z	0	0	0	0	0	0
2014-07-22T18:10:45Z	1	0	0	1	1	1
2014-07-22T18:10:46Z	0	1	1	0	0	0
2014-07-22T18:11:27Z	0	0	0	0	0	0
2014-07-23T15:45:09Z	1	0	0	1	1	1
.....						

Similarly, we calculated the correlation matrix illustrated in Figure 14.

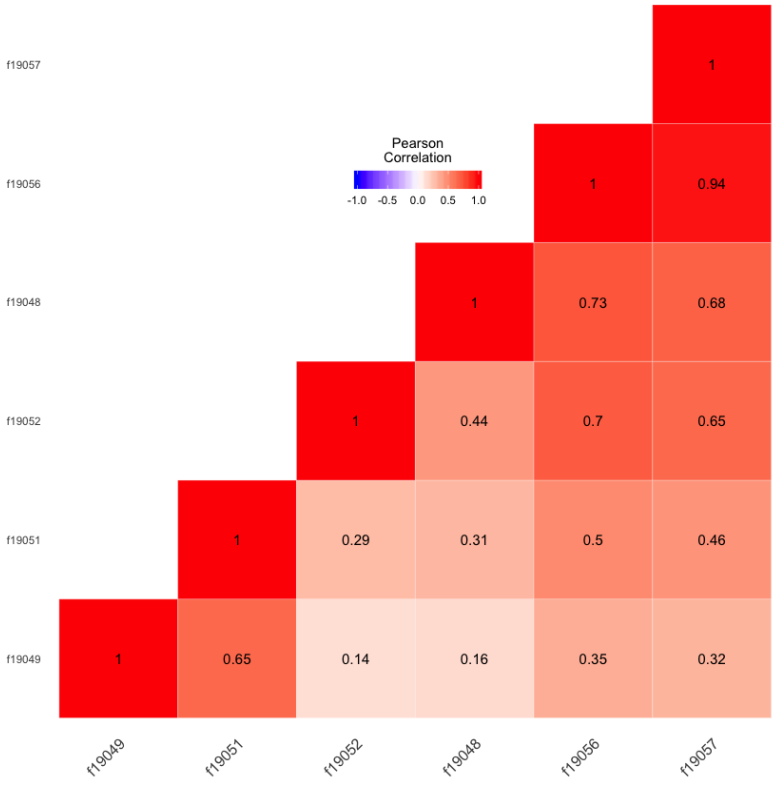


Figure 14 Correlation matrix

Finally, we feed the historical data into neural networks, and obtain the following results.

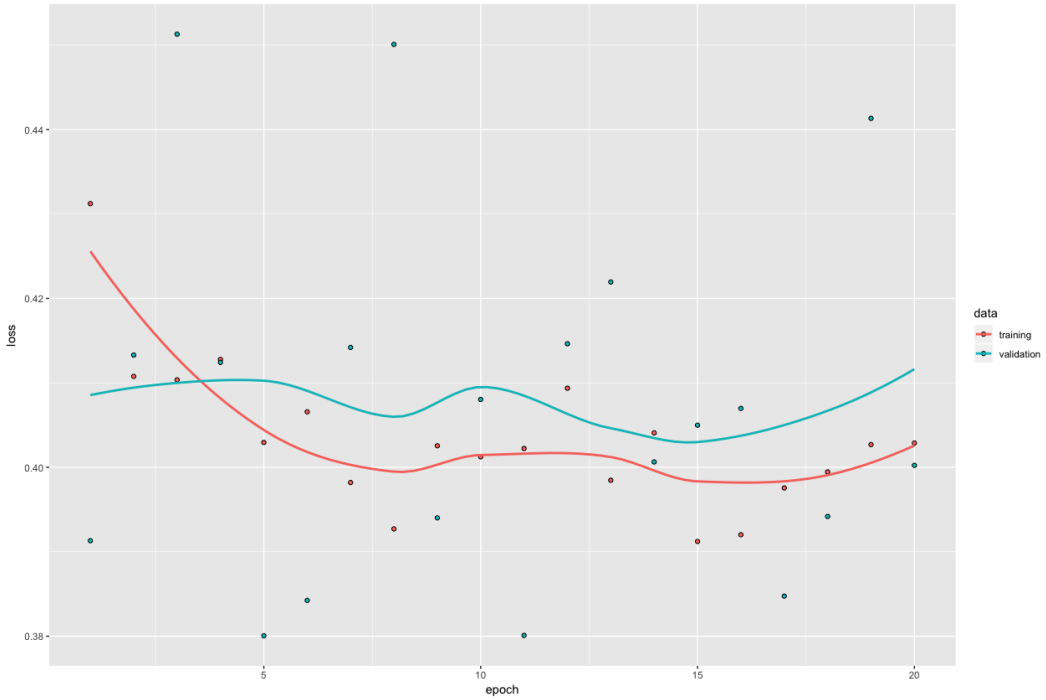


Figure 15: Prediction of failures

Similar to the discussion in our first experiment, it shows the same phenomenon. Compared with two attempts of using neural networks, the second attempt is not as good as we had expected.

Compared with two attempts of using neural networks, the second attempt is not as good as we expected. After having a deep analysis, there are several things which may affect the quality of using neural networks in prediction of failure codes:

- Nature of failure – Sometimes the occurrence of a failure does not have any periodic pattern, thus it is hard to use recurrent neural networks to search.
- Data availability and quality – It is an obvious reason in our case that there are several gaps in the data set.

2.3.4 EARLY WARNING IN MAINTENANCE MANAGEMENT SYSTEM

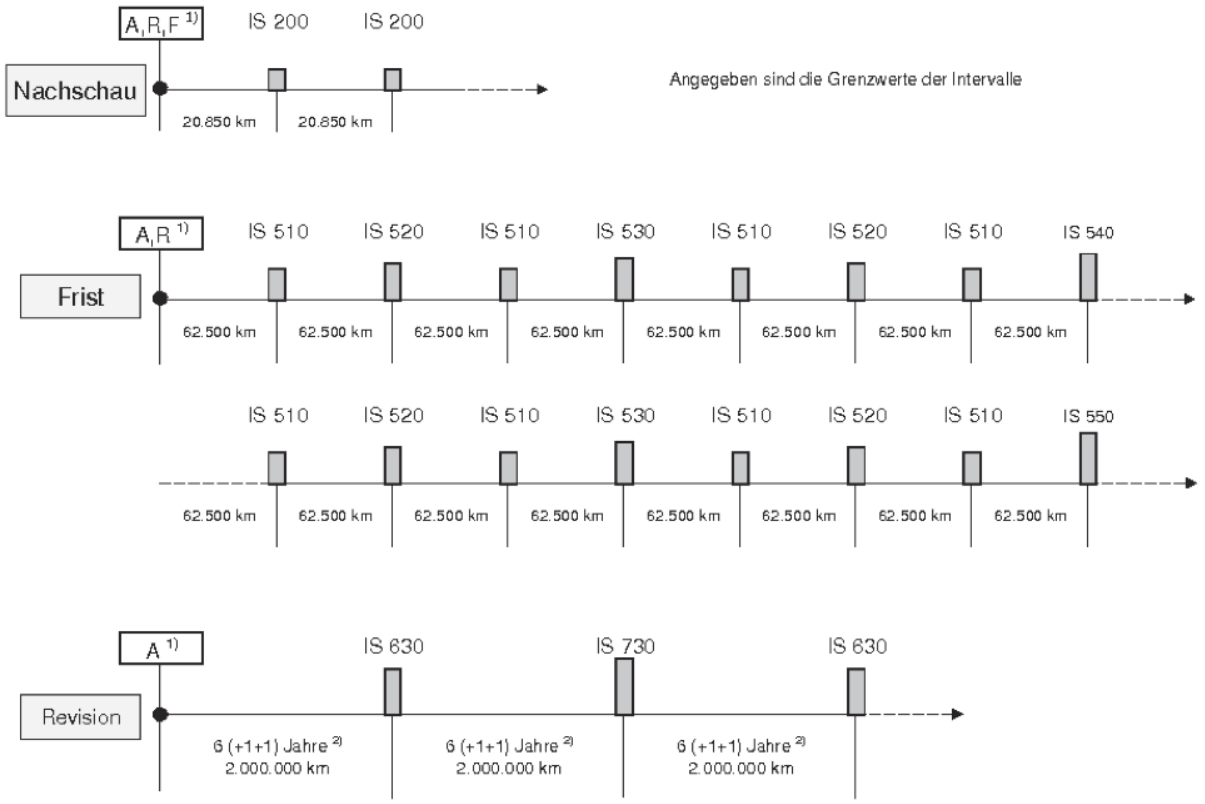


Figure 16: Maintenance programme

The early warning of a potential imminent failure will provide the opportunity to adjust the maintenance programme so that corrective action can be scheduled rather than waiting until a critical failure which would rely on corrective unscheduled maintenance.

Based on the outputs from the prognostic model, the potential occurrence and projected date or mileage of a failure code reaching the specified threshold can be compared to the current maintenance programme, with the aim of adjusting or optimising the maintenance plan where possible. According to the current maintenance programme, as summarised in Figure 16 above, the vehicle is typically inspected every 20,850 km and major (scheduled) services carried out at every 52,500 km. Therefore, the objective of the prognosis techniques is to forecast the failures' occurrence in the next 20,850 km. If the occurrences of several failure codes reach the specified threshold and need to be serviced in a particular service exam (e.g. IS 510, IS520, etc), a new maintenance programme will be created suitable for the needs of that particular fleet.

2.4 CASE STUDY OF HIGH-SPEED EMU

The second case study was conducted in collaboration with Siemens Mobility (GmbH) who are one of partners in the cross-cutting IMPACT2 project. The case study investigated the potential use of the diagnostic data in a condition-based maintenance regime using advanced predictive techniques employing artificial intelligence. The assessment methodology detailed in Figure 4 was also followed for this case and included:

- Diagnostic data structuring and mapping to diagnostic data architecture
- Identification and mapping of code severity based on implied safety risk and maintenance planning within the diagnostic data architecture
- Correlation of maintenance actions in response to 'finally broken' codes from the maintenance provider for the ICE3 fleet
- Observation of data using statistical analysis of diagnostic codes to identify patterns and extract features for use in predictive modelling techniques
- Methods and techniques for future failure event forecasting
- Implementation of the prediction methods for the use in fleet maintenance planning

This case study benefited from two datasets and a data architecture describing the on-train diagnostic data for three systems on the ICE3 fleet; air condition unit, traction system and braking system. The ICE3 or intercity-Express 3 is a family of high-speed EMU trains operated and maintained by DB. The dataset consists of the following information:

1. On-train diagnostic data containing 34 parameters for the ICE3 fleet containing 55 vehicles. The diagnostic data contained diagnostic event information from January 2016 to December 2017.
2. Vehicle depot visits according to 'finally broken' codes with high maintenance response classification linked to 3 depots used to maintain the vehicles by the maintenance provider DB. This dataset includes time the vehicle was in the depot, which 'finally broken' diagnostic code the depot visit was in response to and what mileage the depot recorded when the vehicle was in for maintenance action over the same period in time. The dataset contained maintenance information for 48 of the 55 vehicles in the ICE3 fleet.
3. Finally, the diagnostic data architecture; discussed in more detail in Section 2.4.1. This contains detailed information about each of the approximately 5000 codes the on-train diagnostic system records as failure symptoms or failure events. The dataset contained information about the severity of the diagnostic codes in relation to a maintenance response classification.

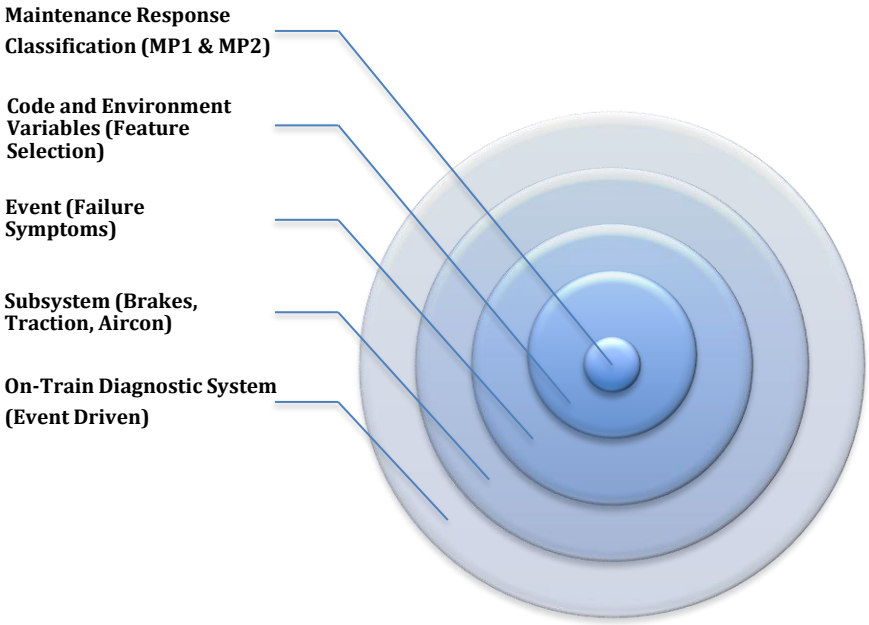


Figure 17: Diagnostic Data Architecture

2.4.1 DIAGNOSTIC DATA ARCHITECTURE

The observed diagnostic data contains a number of parameters that are collected when an event is triggered. The train event recorder attaches 34 of these parameters to each failure event. To develop a usable data structure, a small subset of parameters such as the fleet and vehicle designation, the event failure code, timestamp and other useful parameters were mapped to the diagnostic data architecture shown in Figure 17 above. The combination of event data and mapping of the diagnostic data architecture delivers actionable structure to the event codes based on domain knowledge contained in the diagnostic data architecture itself.

As discussed in the introduction, the structure of the diagnostic information has an architecture designed to be used retrospectively. In normal operation, the events occur while the vehicle is in operation between depots. The data collected in normal operation is specified as a parameter in the diagnostic report and attributed the value of 1 for 'mode of operation' described as operational mode. A value of 2 in 'mode of operation' column indicates that the vehicle is in diagnostic mode, which generally occurs during a maintenance or system testing activity. When the vehicle arrives into the depot the information is downloaded. At this time the data is used to determine the likely cause of failure. For on-train system data, filtering out when the vehicle is in some form of diagnostic mode (e.g. within the depot rather than in operation) allows large quantities of data to be removed from the analysis and used in the subsequent prediction methods.

In the case of the system encountering a threshold breach, a subsystem event triggers recording of specific environment variables. These variables which are continuously monitored by the on-train diagnostic system might include the measurement of some physical conditions of the

system, for example pressure, temperature, vibration etc. so that the information is preserved for future fault diagnosis. This feature is a pattern or programmed behaviour of the diagnostic system that can be employed to determine the severity of the failure code and to forecast future failure events. An example from the braking subsystem is shown in Figure 18 below. The figure shows two high priority codes for the braking system over the same period which results in the vehicle being called in for maintenance actions. The three plots on the left show the cumulative sum of event codes for the 'rear wheelset not turning' and its active duration; the speed of the vehicle and the pressure reading from sensors associated with the failure event are plotted on the bottom plot. The three plots on the right show a failure event code occurring during the same period for the front wheelset. This illustrates the recording of diagnostic failure information.

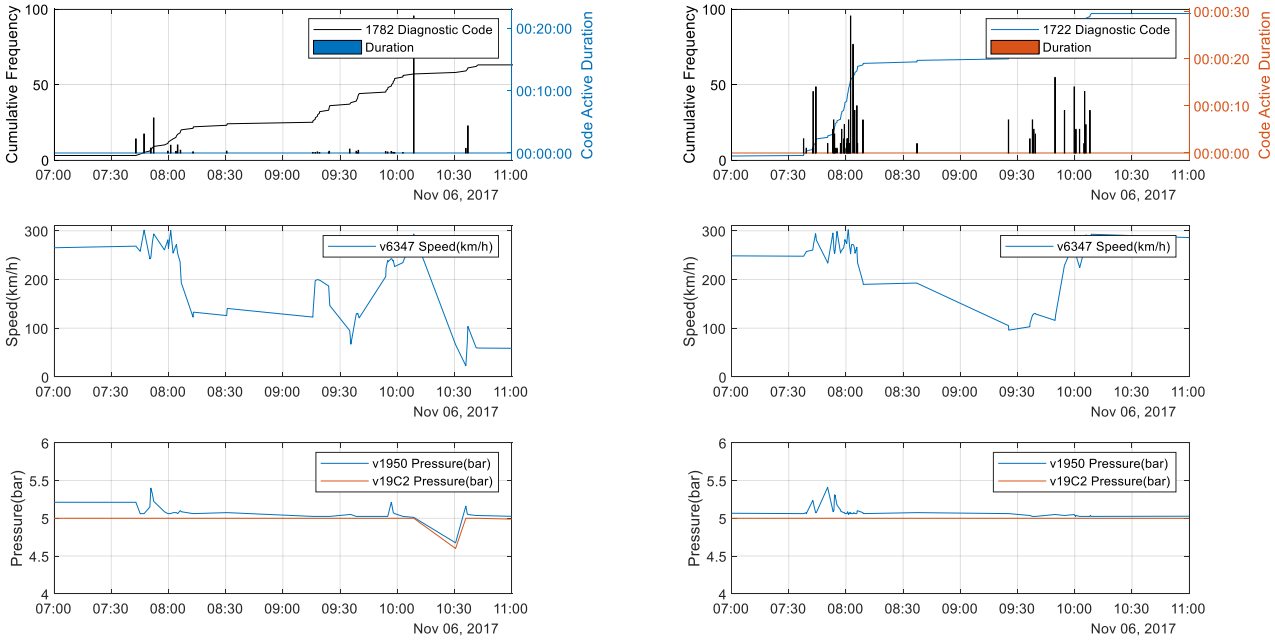


Figure 18: Diagnostic System Behaviour Pattern

The maintenance response classification, at the centre of Figure 17, is assigned to each failure code specified within the data architecture and will typically specified during the design of the diagnostic system. In the diagnostic data architecture, the description of the maintenance response classification the two categories are: Maintenance Premium Standing (MP1) and Maintenance Schedule Volatile (MP2); the first carries a greater warning severity due to its consequences to impact the vehicle operation directly and the second category is a lower severity impacting the maintenance schedule on intermittent basis. This categorisation, along with the environment variable, can be used to prepare for and deal with failures on the arrival of the vehicle during the next depot visit. At the time the event occurs shortly after a period of assessment by the prediction algorithms, if prior warning is delivered to the depot with sufficient information about the failure, it is possible the depot could respond to the event in a proactive manner rather than the current reactive method.

2.4.2 DATA SCIENCE METHODS WITH ENGINEERING DESIGN

Data-driven models can incorporate a range of techniques, from programming to statistical analysis, to machine learning and artificial neural networks. These techniques can be applied to the data collected by the on-train diagnostic system. In order to apply supervised machine learning methods, known failure data is essential (e.g. the occurrence of event codes that have resulted in maintenance actions).

2.4.3 FLEET DEPOT VISITS BY 'FINALLY BROKEN' DIAGNOSTIC CODES

The diagnostic data architecture provides a severity value for each event code based on the importance from a maintenance perspective or safety risk. These codes have a value from 1 to 6; with 1 being most severe and 6 being the least important in terms of maintenance action. The two categories, discussed above, exist in the diagnostic data architecture mapped to each code. This is useful as it can be used to classify the data and attribute a severity to the code and its occurrence in time.

Figure 19 shows the number of depot visits on the y-axis for vehicles in the fleet with maintenance history for the selected 'finally broken' codes. The type of system codes that the vehicles were called in for are shown by the three colours in the legend (e.g. braking system (Bremse), traction system (ASG) and air conditioning unit (Kliminage)). In terms of severity level, mentioned above, these codes have a severity value of greater than or equal to 2.

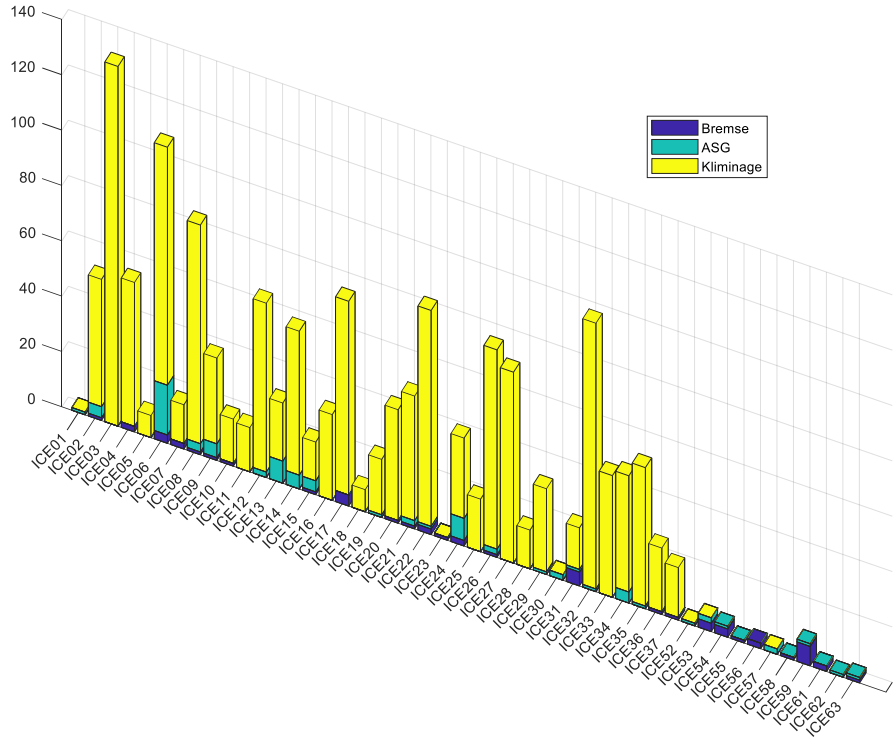


Figure 19: Fleet Depot Maintenance Visits by Diagnostic Code

For the subset of 'finally broken' codes, it can be seen that the air condition (AC) subsystem accounts for the majority of the depot visits with traction and braking systems less affected when the entire fleet is considered. However, some vehicles can be seen to have very low depot visits in total. Provided the sufficient operational data is available, a vehicle with no depot visits can be considered as working within normal operating levels and so can be used in what is known as unsupervised learning methods. In this learning approach ideal condition data, where the system is running within a known condition, is used to teach the system to recognise when the vehicle is running well. Therefore, when the method sees new data that is different to that it was shown previously it recognises it as being abnormal.

In data-driven techniques such as machine and deep learning the quality and quantity of data is of upmost importance. Alongside the diagnostic data and its mapping to a structured architecture, it is important to have failure data that shows good correlation with abnormal system behaviour. The mapped diagnostic data, cross correlated with the depot visits is shown in Figure 20 for the braking, traction and AC subsystems. Approximately 360 event codes are associated with these three subsystems. The dimensionality of the data with this many codes is prohibitively high. Methods to reduce the dimensionality using principle component analysis are common in machine learning methods. They allow predictor variables to be selected that give good responses to training of the classifiers. However, these methods behave well in structured data with constant sampling and continuous monitoring of physical quantities. In contrast, the vehicle diagnostic system records events in a random, non-linear manner dependant on the breach of installed thresholds. Figure 20 shows the results from a detection algorithm designed to extract all mutually coherent codes in

terms of temporal distance for the 'finally broken' code subset. It shows the number of depot visits on the z-axis with frequency of 'finally broken' diagnostic codes, in adjacent columns, correlated in terms occurrence in time. The detection algorithm extracts a reduced code set allowing meaningful pattern recognition and feature selection to be implemented before building a recurrent neural network to predict the occurrence of future failure events. Each instance in the bar chart can be thought of as an opportunity for teaching a neural network the behaviour to recognise in classifying an event for maintenance action.

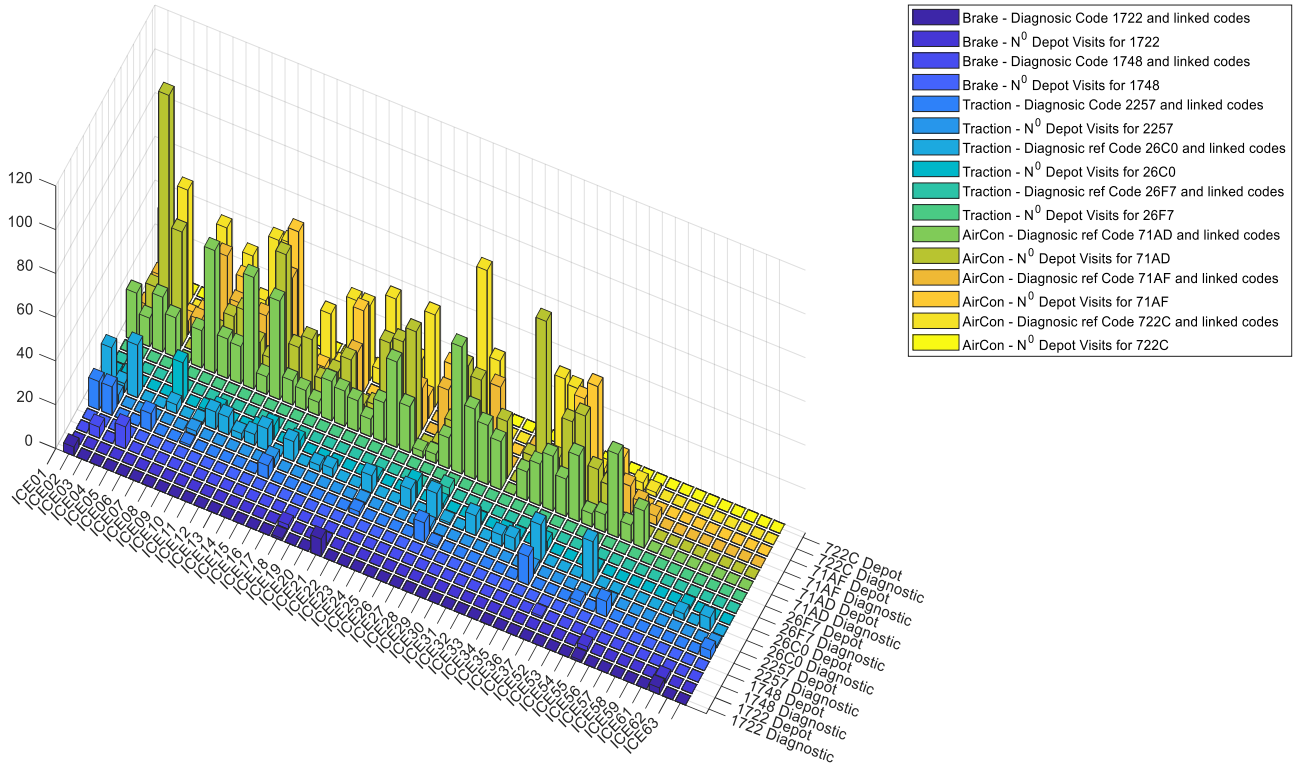


Figure 20: Depot Visits Root Cause Analysis

The number of event codes directly evidenced to depot reported failures are significant and will inform the validation datasets to train the system to recognise, in real-time, impending failures for all the subsystems characterised. It is likely that if more subsystems, preferably all, are included in the analysis it will show greater visibility of the vehicle behaviour and links between subsystem can be discovered.

2.4.4 BUILDING FEED FORWARD NEURAL NETWORKS

Machines can be taught to recognise patterns by feeding forward information about the relationship between the inputs to the outputs through hidden layers of neurons, seen in the left hand illustration in Figure 20. The human brain is a pattern recognition machine built on a huge natural neural networks. Humans can learn from new information by adjusting the strength of neural connections. Critically, the human brain does not observe a pattern and then forget it, it stores it and uses the

information in a feedback network; the right hand illustration, in Figure 21, describes how this behaviour is built up with deeply connected hidden layers.

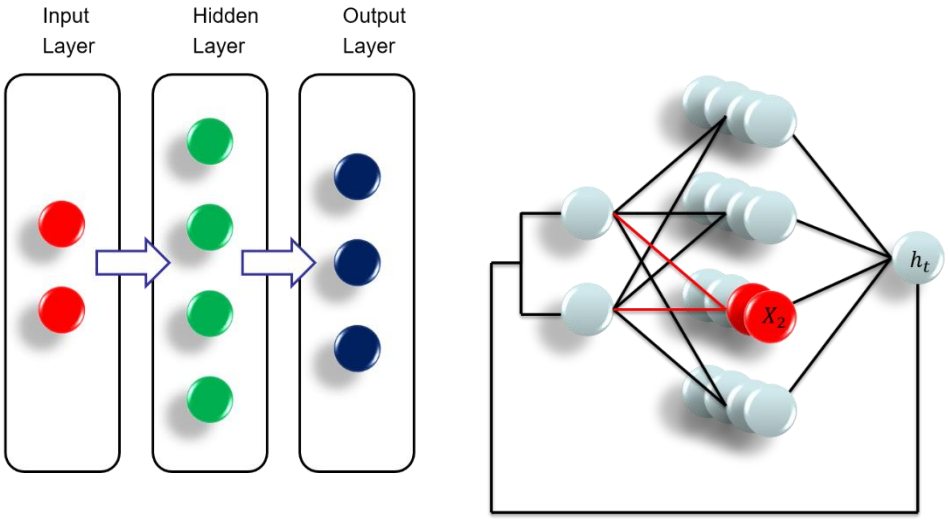


Figure 21: Feed Forward Neural Networks

In an artificial neural network, the connections are assigned numeric values, known as weight to achieve a desired outcome. Each neuron is connected to every neuron in the previous layer. The strength of the connection between two neurons is given by the numeric weight value.

$$X_2 = f(w_{22}x_2 + w_{12}x_1 + b_2)$$

The value passed to the neuron being calculated is done by taking all the value of the neurons in the previous layer and multiplying them by the appropriate weights and summing the results. The sum plus an extra offset, known as the bias is passed to the input of a function known as the transfer function for that layer. The output shown in the equation above is passed to the next neuron. This process is repeated for all the neurons in a layer and then again for the next layer.
The weights, biases and transfer functions determine how inputs are transferred to outputs. Feed forward neural networks are useful for predictive supervised learning problems where the goal is to map a given set of inputs to a given set of outputs.

2.4.5 LONG SHORT-TERM MEMORY NETWORKS

As discussed previously the human brain does not discard learnt information only to relearn from it. It stores the information building on what was learnt and only discards useless information.

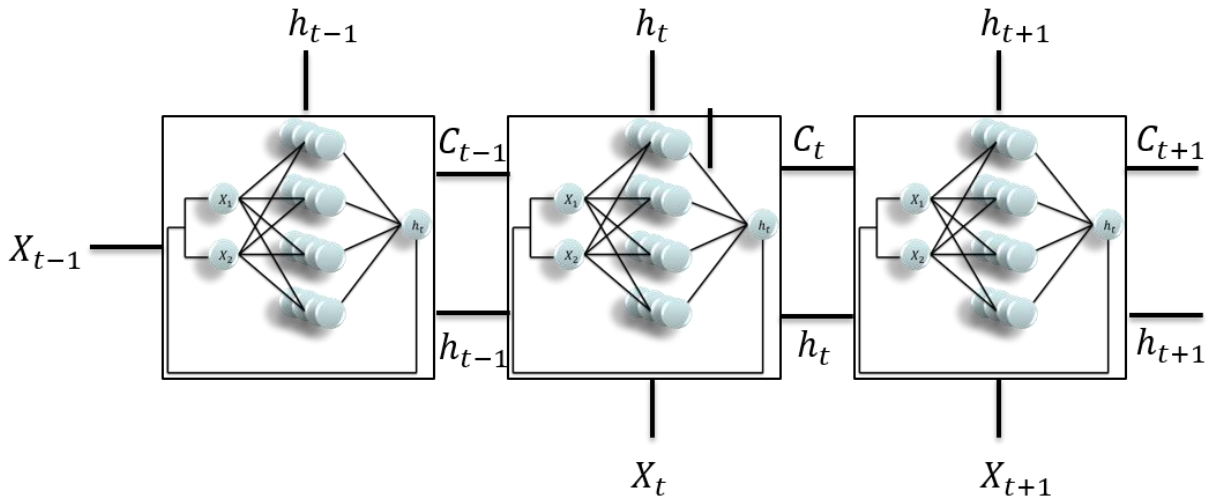


Figure 22: Recurrent Neural Networks

A way of replicating the brains' ability to remember long term dependencies is to use a Long Short-Term Memory (LSTM) network, shown in Figure 22. This is a type of Recurrent Neural Network (RNN). All RNN have this form. In standard RNNs a simple layer within each block, as previously described in Figure 21, can be a simple tanh function, pushing a value from -1 to 1 for the output. In LSTMs, however, this chain like structure is different. Instead of having a single neural network layer, there are four behaving in a distinct manner. The most important element in this structure is the cell state, C_t, the horizontal line running through the top of the figure. The cell state can be understood as a conveyer belt, running through the entire structure with only minor linear interactions. If required, the information being passed from one network to the next flows along unchanged. In an LSTM implementation the RNN has the ability to remove or add information to the cell state, carefully regulated by functions called gates. Gates can let information through and are made up of a sigmoid neural net layer and a pointwise multiplication operation. Simply, the sigmoid layer outputs numbers between zero and one: one, allows all the information to be passed and zero lets nothing through. An LSTM has four such gates to protect and control the cell state by forgetting, updating and outputting the cell (C_t) and hidden (h_t) states.

In an LSTM network the first step is to decide what information is kept and what is discarded from the cell state. This is done through the first sigmoid layer, called the 'forget gate.' It looks at current hidden state h_{t-1} and the next input feature x_t, outputting a value between 0 and 1 for each number in the cell state C_{t-1}. The increments between 0 and 1 representing what is kept and what is discarded. The equation below describes the 'forget gate' layer at time step t.

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right)$$

Where, σ is the sigmoid function, described in the equation:

$$\sigma(x) = (1 + e^{x-1})^{-1}$$

W_f is the learnable weight matrices for the 'forget' layer, h_{t-1} is the current hidden state and x_t is the next input feature at time t.

The next gate is concerned with how the cell and hidden states are updated. This is done in two steps, first a sigmoid layer called the 'input gate' (i_t) is used control which values are updated and the second part which uses a tanh function to create a vector of candidate values (g_t).

$$i_t = \sigma(W_i.[h_{t-1}, x_t\,] + b_i\,)$$
$$g_t = tanh\big(W_g.[t_{t-1}, x_t] + b_g\big)$$

The old cell state C_{t-1} is updated to the new cell state C_t by multiplying by the 'forget gate' f_t to remove information that was discarded earlier. Then input gate and candidate values are concatenated and added to make the new cell state.

$$C_t = f_t * C_{t-1} + i_t * g_t$$

Finally, the output gate (o_t) controls, through filtering of the cell state, what is kept in the hidden state (h_t) as the final output of the LSTM network.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

In the following implementation the feature X_t is cumulative sum of the active high duration for the event code generated by the on-train diagnostic system.

$$\sum_{d_t=1}^{Ef_t} X(d_t)$$

This is calculated from the end of the first active high duration (d_t) up to the point of known failure (Ef_t) where the vehicle was called in for maintenance action.

2.4.6 RESULTS AND FINDINGS

The LSTM RNN is with the historic time series event data (leading up to the maintenance action) obtained from the braking system for a code with maintenance priority level one (high severity). The objective of the LSTM RNN model is to predict the accumulation of activations for the selected code (Active High), which can then be used to support maintenance decisions. A sequence-to-sequence method recalculating the prediction at each iteration is shown in the top left hand plot of Figure 22. When the prediction is based on values at each time step and updated with previous predicted values the error can be higher as seen in the calculated Root Mean Squared Error (RMSE) on the bottom left hand plot, with an RMSE of 19.5.

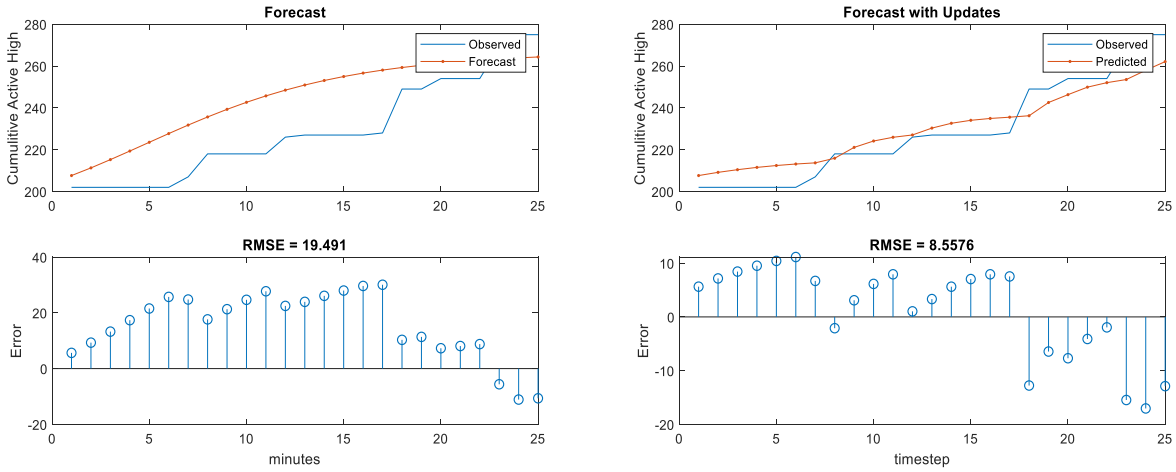


Figure 23: Long Short-Term Memory Network Implementation

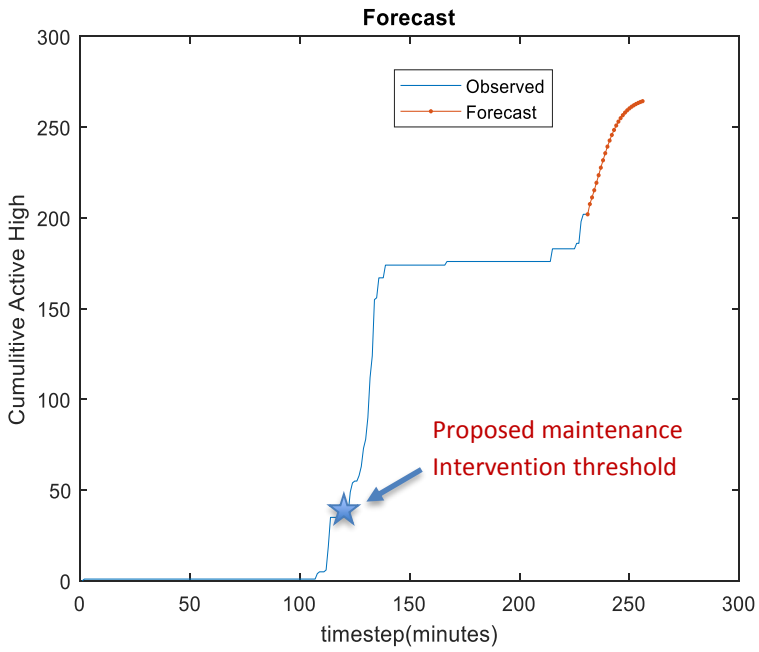


Figure 24: Full Results of LSTM Network Performance

If observed values are available to the LSTM RNN then the RMSE value can be reduced to give a more robust model as can be seen in the right hand plots of Figure 23. Here the forecast is tracking actual accumulated active high time significantly better with an RMSE value of 8.6.

Figure 24 shows the full output implementation from LSTM network for the MP1 brake subsystem code. The graph shows the observed and forecasted values from the RNN. The y-axis shows the accumulated time a code was activated. It can be seen that for this example good correlation between the rate at which the event code is active is predicted within an acceptable error range

compared to the method using unobserved values. With more development and through validation with continuous data it is anticipated the prediction of future events could give pre-warning of failure events some 50-75% earlier to a proposed maintenance intervention threshold as indicated by the blue star in the figure. The prediction threshold can be designed to be the product of accumulation time the code is active and its gradient. Depending on an actual time between depot visits, a reasonable goal for the advanced warning could be between 75 to 100 hours prior warning to prepare the depot to efficiently manage problematic vehicles and more importantly, the management of the entire fleet of vehicles.

2.4.7 PROPOSED IMPLEMENTATION

The implementation of the proposed techniques within a predictive maintenance strategy will require periodic access to data from the on-train diagnostic system. The most reliable method will be to have a microcontroller on-board the train that would have continuous access to the diagnostic data with additional access to the vehicle operational data such as speed, distance and GPS channels. This standalone microcontroller should be equipped with sufficient processing power to replicate all the individual steps to deliver the prediction. It should be capable of communicating advanced warning with sufficient detail about the failure symptoms to ensure an adequate amount of time to enable fleet maintainers to adjust maintenance plans and prepare the vehicles for arrival at the appropriate maintenance depot.

During normal operation, the vehicle may experience a failure event. For example, as can be seen in Figure 25, if this failure event is detected between depot C and depot B, going in an anti-clockwise direction, and the event happens within 24 hours of leaving depot C shown as MP1 in red on the illustration. Furthermore, If the assumed time between depot visits is 3 days and if the high severity failure warning (‘Finally Broken’ MP1 code) is reported by the vehicle, it will be 48 hours away from depot B. With the proposed prediction methods, it is anticipated that informed decision could be made as to the nature of the failure and severity according to the diagnostic data architecture. It is anticipated that this should provide sufficient early warning to modify the current maintenance plan and schedule the vehicle to arrive at the appropriate maintenance depot. The time period before failure should also provide sufficient time to allow the depot (depot B in Figure 25) to prepare the necessary resources, spare parts and equipment etc. to deal with the failure symptoms in a proactive manner to reduce the cost of maintenance, rather than the higher cost associated with unscheduled corrective maintenance.

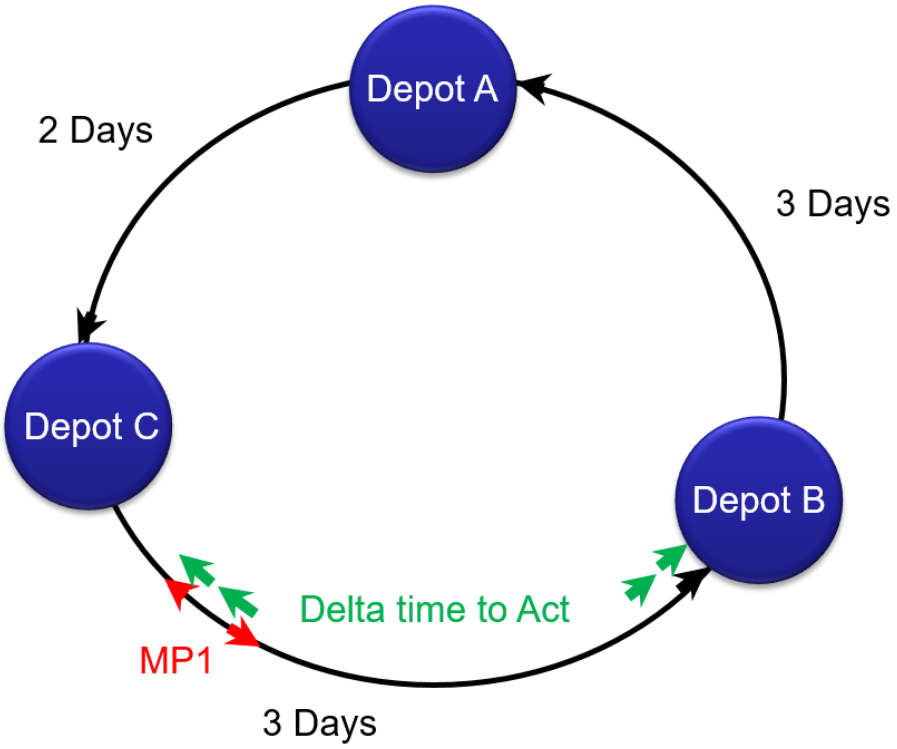


Figure 25: Conceptual Maintenance Planning Implementation

2.5 SUMMARY OF OBSERVATIONS AND RECOMMENDATIONS

The research described in the previous sections has investigated the application of on-train diagnostic data for use in a CBM model for rolling stock. In principle two types of machine learning approaches have been applied to the on-train diagnostic data from the two case studies. The first uses an unsupervised method looking at historical data to extract trends and patterns from the failure symptoms without actual information of failures and maintenance records. The second method is supervised; here failure information is used to teach the neural network the patterns in data as recorded by the on-train diagnostic system. Each recorded failure event, if successfully linked to a 'finally broken' code and maintenance action, can be used as a resource for teaching the model the sequence of events or patterns leading up to a failure which can be used to provide an early warning of an imminent failure and support maintenance planning.

In contrast, it can be seen that the unsupervised method used in case study 1 must infer the failure condition from the data which can be subjective and open to interpretation. Furthermore, a detailed analysis of the network performance using the training and validation datasets presented in the case study reveals that the trained neural network does not respond well in learning the patterns in the historic data due to incomplete data and lack of linked maintenance records. Critically in this type of implementation the lack of data quality, absence of failure data and the process of linking and correlating the dataset to known failure events by maintenance actions has resulting in a poor failure prediction.

In case study 2, using the supervised methodology, a novel data architecture derived from data acquired by the on-train diagnostic system was established which provides a multidimensional view of the operational behaviour for the systems analysed. It also allows the use of statistical methods to be employed to observe trends and patterns in the data more efficiently. Finally, it allows features from that data to be extracted looking at vehicle and fleet wide operation. These can then be used in the development of advanced predictive methods for intelligent scheduling of maintenance.

A valuable addition to the dataset used in case study 2 is the relationship between the failure codes and maintenance actions obtained from the vehicle operator/maintainer, which were used to teach the predictive model to recognise when failure patterns in the diagnostic data were present. This allowed for a more accurate prediction of future failures. The approach is envisaged to yield benefits over an unsupervised methodology where failure information is not employed.

To continue the progress and development of a system for advanced warning and scheduling of future maintenance using on-train diagnostic data, it is recommended that continuous data from a larger range of systems/components from at least one vehicle are analysed to evaluate the use of the prediction methods for early failure warning and maintenance scheduling. Further work should also include the identification of how the prediction techniques could be implemented in the maintenance planning of rolling stock and further validation of the method and outputs.

PART II DECISION SUPPORT BASED ON CONDITION DATA

3. FERTAGUS CASE STUDY

This case study explores the use of condition data to support wheelset maintenance decisions for a Portuguese train operating company – Fertagus. First, Section 3.1 applies Linear Mixed Models (LMMs) to assess the evolution of wear and compares the wear performance to a fleet operating on LUL, while Section 3.2 applies Survival Models (SM) to quantify the probability of damage occurrence. Then, section 3.3 addresses the use of a Markov Decision Process (MDP) approach to derive an optimal maintenance strategy for railway wheelsets. Sections 3.4 and 3.5 explore the application of the prototype of decision support tool for a rolling stock management system, using respectively: i) a tactical maintenance planning and ii) an operational maintenance scheduling. Finally, Section 3.6 provides a discussion on the uncertainty associated with different inspection devices (e.g. laser inspection device) potentially used in the CBM model.

3.1 STATISTICAL MODELLING OF WHEELSET WEAR

This section explores the use of statistical models to assess the evolution of wear trajectories of railway wheelsets. It provides insight into the process of wheelset degradation and their usual maintenance procedures. Using a quantitative basis of data from a fleet of modern EMU trains from a Portuguese train operating company, different model specifications for the wheelsets' wear evolution are compared using Linear Mixed Models (LMMs). The wear trajectory is assessed by the evolution of the wheel tread diameter, the flange thickness, the flange height and the flange slope. The variability in the data was associated with several factors, such as the month of measurement, the unit vehicle or the vehicle type, and their influence on the wear trajectories was also analysed. From the observation of the results obtained, it was possible to conclude that the wheel hardness can have an influence on the wheelset degradation trajectory. Finally, the statistical patterns found seem to be consistent with other train fleets.

- **Brief background**

This subsection focuses on the wheel condition, particularly on the statistical modelling of the deterioration processes involved in the wheel wear evolution. It provides a quantitative basis, based on a sample collected from a Portuguese train operating company, which may provide a better understanding of the needs in wheelset maintenance processes (renewal, preventive and corrective). It also supports the identification of the main factors that explain the variability in wear predictions. Finally, it also corroborates the theory that the models and variables here adopted can be applied to any fleet of vehicles revealing similar patterns and behaviours (Andrade and Stow 2016).

Statistical approaches to study the wear behaviour in the degradation of railway wheelsets are more commonly found in the analysis of physical quantities, such as vertical wheel loads, residual

stresses, longitudinal or transverse contact stresses, rather than geometric parameters (Pombo et al. 2011a, Hossein et al. 2015). In fact, most of the studies do not cover the probabilistic issues in their modelling, mainly when irregularities can be considered continuously distributed along the track. In these cases, approaches based on stochastic process theory are more appropriate (Iwnicki 2006). A few studies investigated degradation data from the wheel profile to estimate failure distributions and associated reliability (Freitas et al. 2009, Asplund et al. 2016). Moreover, Lin and Asplund (2014) used Weibull models to estimate lifetime data for a sample of locomotive wheels. Wang et al. (2015) used a data-driven model to optimize the wheel reprofiling strategy, aiming to extend the life cycle of metro wheels. Recently, from the perspective of reducing life cycle cost and managing wheelset maintenance activities, different Markovian approaches were conducted to optimize the reprofiling policy for train wheels, by modelling distinct variables to identify degradation states (Jiang et al. 2017, Braga and Andrade 2018, Mingcheng et al. 2018).

Notwithstanding, none of the above presented statistical studies used Linear Mixed Models (LMMs) in literature. The one relevant was a previous research work of Andrade and Stow (2016), whose analysis was further used as a basis for a new wheelset maintenance strategy, called ‘economic tyre turning’ in Andrade and Stow (2017a). This present paper follows the study and methods used in Andrade and Stow (2016) and tries to give a clear answer to a few main topics left open for further research, as to whether or not the statistical patterns found are consistent in other train fleets. Therefore, it tries to validate the statement that these LMMs can be applied to any fleet of vehicles with consistent results. Secondly, the present research work also introduces a new important variable - the flange slope (qR) - which is in line with what is proposed in Asplund et al. (2016), due to its importance on the control of the degradation and damage of the wheel profile. Finally, this section also assesses the influence of the wheel hardness in the wheelsets degradation trajectories. The sample analysed – from a case study on the fleet of a Portuguese train operating company – went through a big renewal program in its train fleet. Every wheelset was renewed by a new one, with wheels with different hardness. Therefore, this statistical analysis also makes the distinction between this two operating cycles, considering its influence in the wheelset wear trajectory.

A railway wheelset is a component that consists of two wheels linked by a rigid axle, allowing the motion to the vehicle when rolling over surfaces (rails), as depicted in Figure 26.

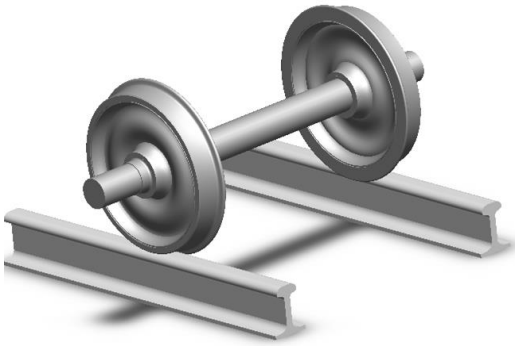


Figure 26: Railway wheelset and rails.

For safe operation of the wheelsets, it is essential to guarantee that both the axle and the wheels are not damaged and are within the dimensional safety specifications. Otherwise, both the axle or the wheels have to be reprofiled or replaced by new ones if necessary, which will be a request of maintenance.

To control the level of degradation of the wheels it is necessary to periodically assess some geometric variables from the wheel tread profile (Figure 27) which are measured relatively to three fixed measurements (a, b, c) and from a tread datum position point (T). If these variables are beyond the safety limits, the wheelset has to be reprofiled or replaced.

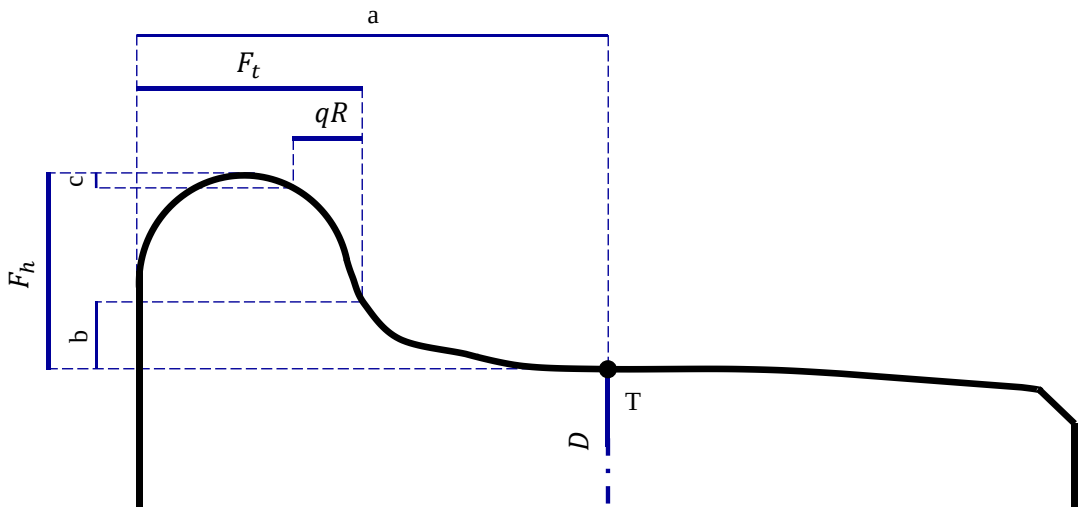


Figure 27: Wheel diameter (D), flange height (Fh), flange thickness (Ft) and flange slope (qR).

To illustrate schematically typical wear trajectories of railway wheels and their maintenance, Figure 28 is provided, using the wheel diameter (D) as the main indicator. Continuous blue lines represent the actual deterioration process of the wheels on the wear trajectory. Note that, for simpler understanding, it is considered that wheels wear at a constant rate, i.e. the continuous blue lines have the same slope in the graph. The blue dotted lines represent the impact in D due to the maintenance actions performed.

Railway wheels are in service starting from an initial diameter (Di), when they are new (green squares in Figure 28), until the diameter reaches the scrap diameter (Ds), beyond which it is not safe to continue operating, the vehicle must be removed from service and the wheelset replaced (renewal). Moreover, there are running profile limits for the flange height (Fh_lim), flange thickness (Ft_lim) and flange slope (qR_lim). To avoid the wheelsets reaching these case limits and to prevent them from other non-detected problems, preventive maintenance (turning) is carried out with a certain kilometre interval (or mileage interval).

Typically, train operating companies do this type of maintenance after an established number of unit kilometres, since the last maintenance operation (turning or renewal). This is the reason why, in Figure 28, preceding each preventive maintenance (yellow squares), there is the same wear trajectory, i.e. the same line slope a line with a same length and unit kilometres between maintenance operations. In different circumstances, if the wheels are found beyond the limits before the next preventive maintenance and if the wheels have not reached yet their scrap diameter, the wheels must be reprofiled (turning) to restore the geometric parameters (Fh, Ft, qR) to safer values, this is what is called a corrective maintenance action represented in Figure 28 with the orange square. Moreover, if the wheelset is found at any time with damage, it must also go to the wheel turning lathe for corrective maintenance (red square in Figure 28). It is assumed that each kilometre interval between maintenance operations (turning or renewal) is a variable called kilometres since last turning or renewal operation (K).

Each time a wheel goes to a wheel lathe, it undergoes a diameter loss due to turning (ΔDT) which can be higher or lower, depending on the maintenance type action being taken and the specific situation (e.g. presence of damage, as wheels flats, cavities or Rolling Contact Fatigue). In a situation of preventive maintenance, it is expected that the wheel goes through the smallest loss of diameter. On the other hand, to correct damaged wheels, it takes a big diameter loss, shortening significantly the wheel life cycle (Pombo et al. 2011b). In fact, this last situation can be seen in Figure 28. It is possible to distinguish two distinct wheel life cycles: the first one that includes K1, K2, K3, K4 and a second one that includes K5, K6, K7, K8. The first wheel cycle had an extended life because it went only through regular preventive maintenances. By comparing the cumulative kilometres since turning of the first cycle (K1 + K2 + K3 + K4) with the second one (K5 + K6 + K7 + K8), it is clear that the latter had a much lower span life. This is not only because of the corrective maintenances that this wheel went through, but even more due to the damage correction (red square) that, in Figure 28, occurred at a time of a lower diameter. In fact, Figure 28 goes in line with practical observations reported in the past, in which there is a greater probability of damage occurrence in smaller diameters (Molyneux-Berry and Bevan 2012).

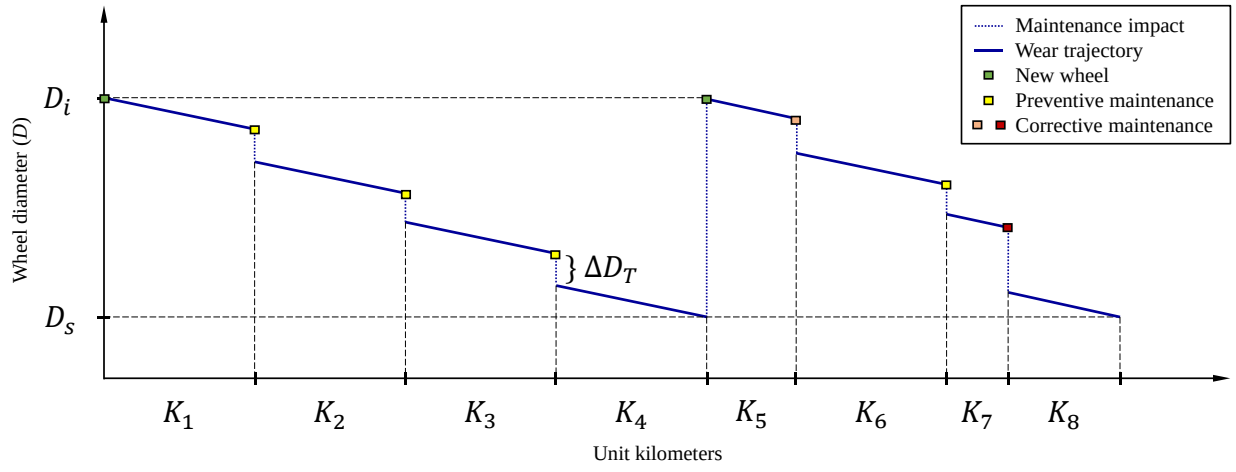


Figure 28: Schematic wheel maintenance trajectories with wheel diameters and the unit kilometres.

The use of LMMs in statistical modelling of wheel degradation can be advantageous to infer about the dependence between different variables in the wheel wear evolution, to provide straightforward mechanisms to control for the variability within and between different groups of the wheelset position and technical specifications (Molyneux-Berry and Bevan 2012, Ferreira et al. 2012). LMMs are linear models that both monitor the fixed effects of different controlling variables (Xiβ) in the expected mean of the dependent variable and the random effects associated with some factor or group (Zibi). According to Galecki and Burzykowski (2013), and for a single grouping level, LLMs can be formulated as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i + \varepsilon_i$$

Where y_i is the dependent variable for the group i, X_i is the designed matrix for that group i, β is the slope parameter, ε_i is the residual error for group i, Z_i is the matrix of covariates for group i and $\boldsymbol{b}_i$ is its corresponding random effect. It is assumed that the random effects (bi) and the residual errors (εi) follow normal distributions with zero mean and covariance matrices of D and $\mathcal{R}$i, with bi ⊥ εi

$$\boldsymbol{b}_i \sim N(\mathbf{0}, D)$$

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Re_i)$$

Both terms bi and εi are considered independent for the same group i and between different groups. The covariance matrices are specified with an unknown scale parameter σ^2 as follows:

$$D = \sigma^2\boldsymbol{D}$$

$$\Re_i = \sigma^2\boldsymbol{R}_i$$

Note that there are a few additional constraints that have to be made on the matrices D and Ri - multiples of the identity matrix - to guarantee identifiability (Galecki and Burzykowski 2013). All these statistical models were estimated using the 'lme4' package for the R software (Bates 2010, 2018).

- **Application to Fertagus case study:**

Fertagus is a Portuguese train operating company, which is part of *Grupo Barraqueiro*, and became the first private train operator to guarantee the commercial concession of a railway line in Portugal. This company is responsible for ensuring the suburban passenger transportation between 14 railway stations from Roma-Areeiro (Lisbon) to Setúbal.

The data analysed comes from wheelset turning maintenance operations, of a fleet of 18 EMU trains of a single type or class, between October 2000 up to June 2015 (i.e. a 16-year interval). Each unit has four vehicles and each vehicle has eight wheels (i.e. four wheelsets). Figure 29 provides a schematic representation of a four-car unit.

The process of data extraction took several visits to the maintenance yards (situated in Coina) and the access to their maintenance actions archive. The information on the geometric parameters that control the wear evolution of the wheelsets was saved in paper format, since it comes directly printed from the CNC (Computer Numerical Control) machine of the under-floor wheel lathe used each time a reprofiling maintenance action occurs. Because of that, before the data was able to be treated, it was necessary to use Computer Vision procedures, to convert the numerical information in the turning sheets into digital format. The turning sheets have information of the wheel profile degradation measurements - i.e. wheel diameter (D), flange height (Fh), flange thickness (Ft) and flange slope (qR) - pre and post-turning, in preventive and corrective maintenance actions.

The process of wheelset turning can be described as follows: the vehicle arrives at the under-floor wheel lathe (which is from the Spanish train manufacturer Talgo) and the technician starts by fixing the wheelset to the turning machine, then, the CNC machine is calibrated relatively to the wheelset position and, finally, the turning starts. By the time of the turning, the technician has also to guarantee that there is no significant difference in diameters between wheels of the same wheelset, wheels of the same bogie and wheels of different bogies. The process of corrective maintenance actions takes more time than preventive maintenance actions, and the influence of the technician experience and sensitivity is more predominant. Regarding the technician influence in wheel wear maintenance operations, Société Nationale des Chemins de fer Français (SNCF) attempts to combine quantitative data with perceptions and experience of the wheel maintainers (thus, adding a subjective dimension to risk assessment) in order to tackle organizational issues with multiple decision makers and multiple criteria (Tea 2012). Another contribution towards the incorporation of the technician experience focused on the variability between the different wheel lathe operators (Andrade and Stow 2017b).

Regarding other case studies involving Talgo turning lathe machines, Talgo developed a maintenance program called Total Logistic Care that keeps the flange thickness within an ‘optimal’ range of operation, instead of waiting until the wheel is out of the specifications (Pascual and Marcos 2004).

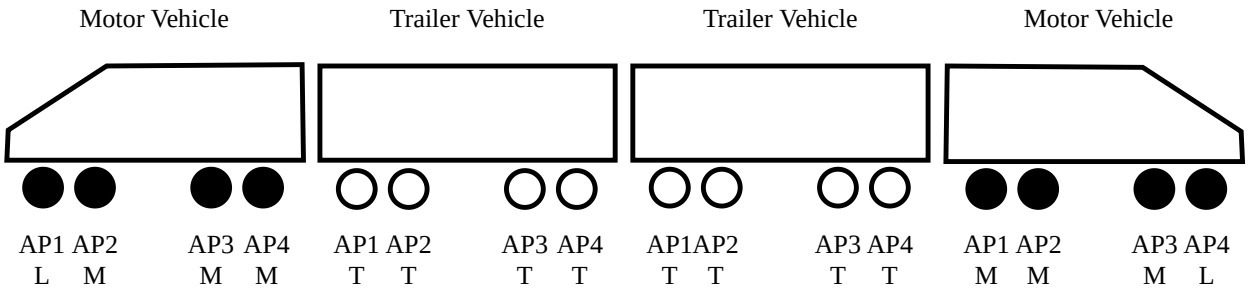


Figure 29: Schematic representation of a four-car unit with four axle positions (AP1 - AP4).

Fertagus went through a big revision in their train fleet. Every wheelset was renewed by a new one, with wheels made of different materials with different hardness. Each train unit changed every

wheel at once at a certain time between 2011 and 2013. Assuming that wheels with different hardness will have different wear trajectories, the following analysis using LMMs splits the two operating wheel cycles:

- Cycle 1 (C1): wheels with the material of type 1;
- Cycle 2 (C2): wheels with the material of type 2.

The wheel database used in this research contained the following information: unit number (18 units), unit running kilometres (cumulative kilometres), vehicle type (motor or trailer), date (cumulative months), wheelset position (16 positions), tread diameter (pre and post-turning), flange height (pre and post-turning), flange thickness (pre and post-turning) and flange slope (pre and post-turning). The CNC machine, from where the measurements of the wheel profiles were withdrawn, has the precision of ±0.05 mm for the wheel diameter and ±0.1 mm for the remaining geometric indicators (the flange height, flange thickness and flange slope). Table 1 provides an overview of the variables used in the analysis, their description, type, some statistics (mean, minimum and maximum), as well as the values precisions.

Table 1: Variables, their description, type, some statistics and precision.

Variables	Description	Type	Mean	Min	Max	Precision
$\lvert\Delta D\rvert$	Diameter loss due to wear [mm]	Continuous	5.90	0.05	18.50	±0.05
ΔF_h	Change in the flange height due to wear [mm]	Continuous	2.6	-0.4	8.0	±0.1
ΔFt	The change in the flange thickness due to wear [mm]	Continuous	0.6	-8.2	6.6	±0.1
ΔqR	Change in the flange slope due to wear [mm]	Continuous	0.7	-5.5	4.2	±0.1
K	Kilometers since last turning/renewal [km]	Continuous	161982	5000	343662	±1
D	Tread diameter pre-turning [mm]	Continuous	866.10	797.55	924.70	±0.05
F_h	Flange height pre-turning [mm]	Continuous	30.8	27.8	38.0	±0.1
Ft	Flange thickness pre-turning [mm]	Continuous	31.6	16.1	36.4	±0.1
qR	Flange slope pre-turning [mm]	Continuous	11.2	4.6	15.7	±0.1
W	Wheelset type (3 types: motor, trailer, motor leader)	Nominal	–	–	–	–
H	Hardness (2 types: C1, C2)	Nominal	–	–	–	–
U	Unit number (18 units)	Nominal	–	–	–	–
V	Vehicle type (2 types: motor vehicle, trailer vehicle)	Nominal	–	–	–	–
M	Month of measurement (cumulative)	Nominal	–	–	–	–

On the wear trajectory, it is necessary to study the variables that assess the evolution of the geometrical measures of the wheel profile, which are the change in the tread diameter due to wear (ΔD), the change in the flange height due to wear (ΔFh), the change in the flange thickness due to wear (ΔFt) and the change in the flange slope due to wear (ΔqR). Going back to Figure 28, to the case of the wheel tread diameter (D), the change in diameter due to wear (ΔD) is the difference between the final and the initial wheel diameter for each graph segment in continuous blue lines (i.e. each wear period). Similarly, it is possible to extend this difference to the remaining wheel profile measurements and define the quantities ΔFh, ΔFt and ΔqR. Note that, in this paper, the

change in the tread diameter due to wear (ΔD) is represented in its absolute value, in a variable called diameter loss due to wear (|ΔD|).

If plotted several observations in the diameter loss due to turning (|ΔD|), the change in the flange height due to wear (ΔFh), the change in the flange thickness due to wear (ΔFt) and the change in the flange slope due to wear (ΔqR), respectively Figures 30 – 33, associated with the kilometres since last turning/renewal (K). It is possible to see significant level of unexplained variability (Table 2), i.e. variability that is not explained by the variation in the kilometres since last turning/renewal.

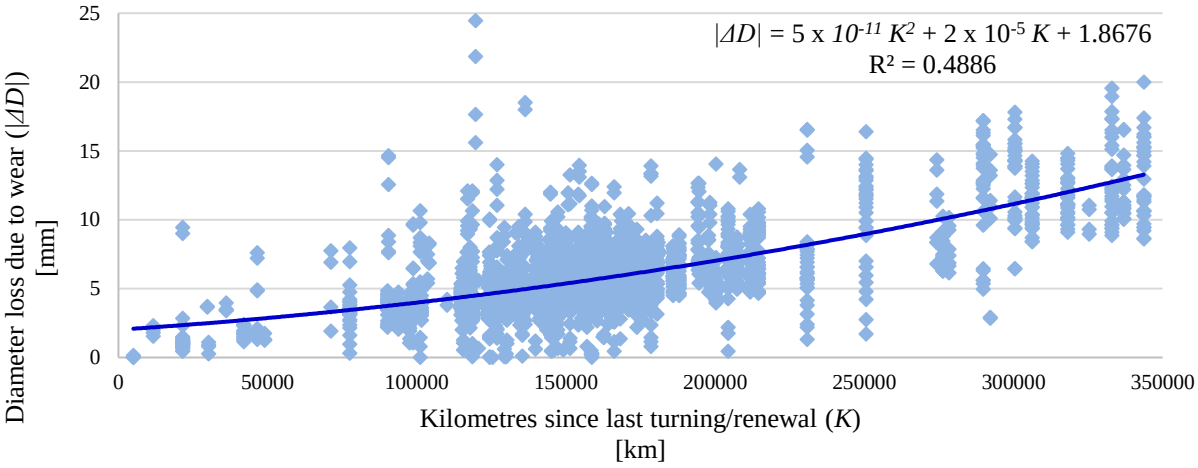


Figure 30: Diameter loss due to wear with the kilometres since turning/renewal.

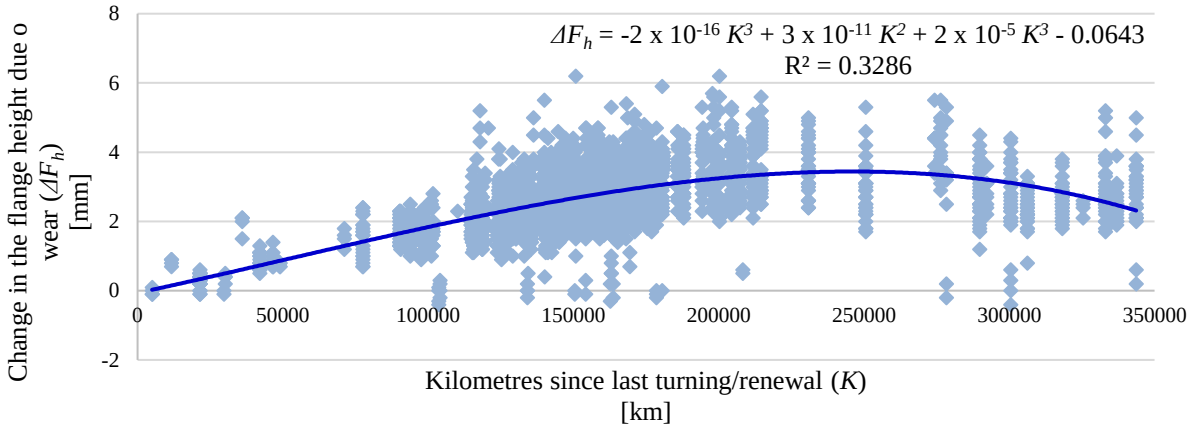


Figure 31: Change in the flange height due to wear with the kilometres since turning/renewal.

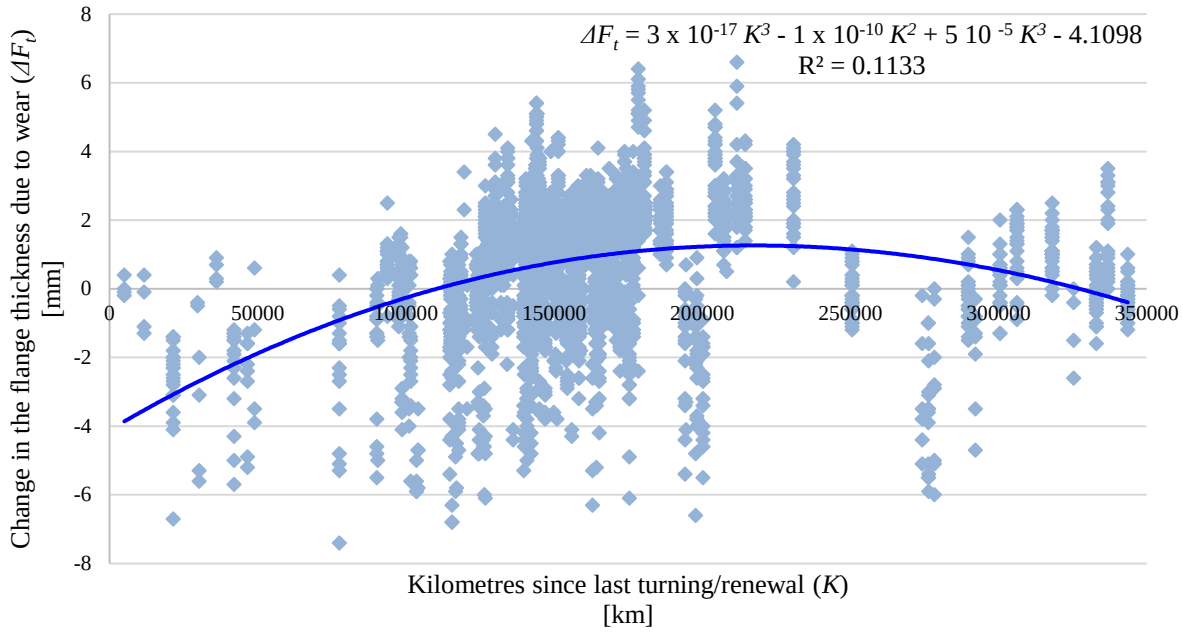


Figure 32: Change in the flange thickness due to wear with the kilometres since turning/renewal.

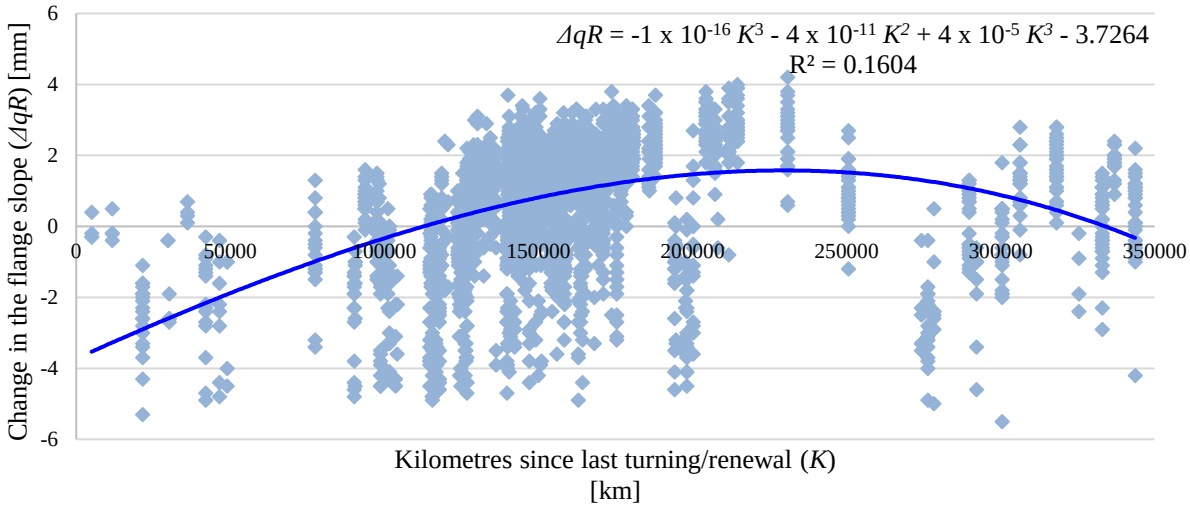


Figure 33: Change in the flange slope due to wear with the kilometres since turning/renewal.

Table 2. Coefficient of determination (R^2) in the wheel profile measurements.

	R^2
ΔD	0.4886
ΔF_h	0.3286
ΔF_t	0.1133
ΔqR	0.1604

However, this variability may be explained by several factors, such as the unit number, the vehicle type and the month of measurement from Table 1. In fact, an LMM concept can handle these factors treating them as random effects in its modelling. There are two ways of modelling random effects with multiple groups: considering them as crossed random effects or nested random effects. For example, modelling the wheelset degradation from a wheelset in a given vehicle, it is possible to consider random effects in wheelset position 'nested' within each vehicle type, or not consider the random effect of wheelset position within each vehicle type and instead model these random effects in a crossed manner. In line with Andrade and Stow (2016), only crossed random effects were used because no statistically significant increase in information is found when nested random effects are considered.

From the 6556 wheel profiles measured, different LMMs were specified for the dependent variables that assess the wheelset's degradation trajectory:

i. The diameter loss due to wear – $|\Delta D|$;
ii. The change in the flange height due to wear – ΔFh;
iii. The change in the flange thickness due to wear – ΔFt;
iv. The change in the flange slope due to wear – ΔqR.

Table 3 compiles and identifies all the fixed effects, random effects and variance structure for each dependent variable in the models here specified (M0 – M4b). The models are associated with the fixed effects of the kilometres since turning/renewal (K), the wheelset type (W) and the wheel hardness (H) - parameters that are known to be strongly related with the wheel degradation trajectory and that are 'fixed' factors inherent to a wheelset, at any time. Then, random effects are added: the month of measurement (M), the unit number (U) and then vehicle type (V).

Models M0 are the simplest ones only with an intercept and a slope parameter, considering only the kilometres since last turning/renewal (K) as the explaining fixed effect variable, since it is the most important fixed effect here analysed. Models M1 are the reference models which consider all the fixed effects here analysed for the dependent variables, but do not take into account random effects.

Note that, some of the models explored were specified in the same way as in Andrade and Stow (2016), i.e. using kilometres since turning/renewal (K) as an explaining variable with two terms: a linear and a quadratic term (M1a – M4a), and others with three terms: a linear, a quadratic and a cubic term (M1b – M4b).

For the models with random effects (M2 – M4), the number of random factors increase, i.e. to the month of measurement (M) in the M2, the unit number (U) was added in the M3, followed by the addition of the vehicle type (V) in M4. This specific adding order was followed by Andrade and Stow (2016), since in their case study this would better identify which random factors added more variability around the expected mean (i.e. controlling for different values for the fixed effects). In

terms of variance structure, the variances for the different groups within each different random effect factor are all considered the same – VC.

Table 3: Linear Mixed Models explored for each dependent variable with fixed effects, random effects and variance structure.

Dependent variable	Models	Fixed effects	Random effects	Variance structure
$\lvert\Delta D\rvert$	M0	1, *K*	–	–
	M1a	1, *K*, K^2, *W*, *H*	–	–
	M2a	1, *K*, K^2, *W*, *H*	*M*	VC
	M3a	1, *K*, K^2, *W*, *H*	*M*, *U*	VC, VC
	M4a	1, *K*, K^2, *W*, *H*	*M*, *U*, *V*	VC, VC, VC
ΔF_h	M0	1, *K*	–	–
	M1b	1, *K*, K^2, K^3, *W*, *H*	–	–
	M2b	1, *K*, K^2, K^3, *W*, *H*	*M*	VC
	M3b	1, *K*, K^2, K^3, *W*, *H*	*M*, *U*	VC, VC
	M4b	1, *K*, K^2, K^3, *W*, *H*	*M*, *U*, *V*	VC, VC, VC
ΔF_t	M0	1, *K*	–	–
	M1b	1, *K*, K^2, K^3, *W*, *H*	–	–
	M2b	1, *K*, K^2, K^3, *W*, *H*	*M*	VC
	M3b	1, *K*, K^2, K^3, *W*, *H*	*M*, *U*	VC, VC
	M4b	1, *K*, K^2, K^3, *W*, *H*	*M*, *U*, *V*	VC, VC, VC
ΔqR	M0	1, *K*	–	–
	M1b	1, *K*, K^2, K^3, *W*, *H*	–	–
	M2b	1, *K*, K^2, K^3, *W*, *H*	*M*	VC
	M3b	1, *K*, K^2, K^3, *W*, *H*	*M*, *U*	VC, VC
	M4b	1, *K*, K^2, K^3, *W*, *H*	*M*, *U*, *V*	VC, VC, VC

For instance, a specific second-degree polynomial, that could model any dependent variable as |ΔD|, ΔFh, ΔFt or ΔqR, would result in the following expression

$$\boldsymbol{y}_{mui} = \boldsymbol{\beta_0} + \boldsymbol{\beta_K K}_{mui} + \boldsymbol{\beta}_{\boldsymbol{K}^2}\boldsymbol{K}^2_{mui} + \boldsymbol{\beta_W W} + \boldsymbol{\beta_H H} + \boldsymbol{b}_{0m} + \boldsymbol{b}_{0u} + \boldsymbol{\varepsilon}_{mui}$$

Considered for the fixed effects on the kilometres since turning/renewal (K), the wheelset type (W) and the wheelset hardness (H), where m indexes the month of measurement, u indexes the train unit, i indexes the individual measurement of the wheel, $\boldsymbol{b}_{0m}$ and $\boldsymbol{b}_{0u}$ are crossed random effects and $\boldsymbol{\varepsilon}_{mui}$ is the traditional normally distributed random error.

In the analysis followed hereinafter, the Akaike information criterion (AIC) is used solely to compare models with different fixed effects and without random effects. On the other hand, the restricted maximum likelihood (REML) criterion, namely a 'goodness of feet' measure: the −2 restricted log likelihood, is used to compare models with the same fixed effects but different random effects. The reason why the model comparison is conducted using the restricted maximum likelihood (REML) criterion is due to the 'lme4' package fits the model using that same criterion (Bates et al. 2014).

For a deeper discussion on the use of different criteria in model comparison in LMM, see Müller et al. (2013), namely on the lack of consensus on how to approach model selection in LMM.

i. The diameter loss due to wear – |ΔD|

The first dependent variable that needs to be modelled is the diameter loss due to wear (|ΔD|). As explained before in Figure 30, there is a lot of unexplained variability around the second-order polynomial describing the evolution of the diameter loss due to wear with the kilometres since last turning/renewal. This variability is then explored again through LLMs, comparing the different specifications in Table 3 for the models M0–M4a. Table 4 provides the REML estimates for the parameters of the models explored. Note that, all the coefficients are statistically significant at the 5% significance level for all fixed effects. Comparing the variances with the total variance (σ2 + dM + dU + dV = 3.804), it is possible to find out that the measurement noise still captures 85.4%, the factor month of measurement (M) captures 12.3%, the factor unit (U) captures 2.2% and finally the factor vehicle (V) captures 0.1% of the total variance.

ii. The change in the flange height – ΔFh

The second dependent variable being modelled is the change in the flange height due to wear (ΔFh). As explained before in Figure 31, there is a lot of unexplained variability around the third-order polynomial describing the evolution of the flange height due to wear with the kilometres since last turning/renewal. This variability is then explored again through LLMs, comparing the different specifications in Table 3 for the models M0–M4b. Table 5 provides the REML estimates for the parameters of the models explored. Note that, all the coefficients are statistically significant at the 5% significance level for all fixed effects. Comparing the variances with the total variance (σ2 + dM + dU + dV = 0.8054), it is possible to find out that the measurement noise still captures 52.4%, the factor month of measurement (M) captures 17.9%, the factor unit (U) captures 1.3% and finally the factor vehicle (V) captures 28.4% of the total variance.

Table 4: Restricted maximum likelihood estimates for the parameters of models M0 – M4a for the dependent variable change in the tread diameter (ΔD).

		ΔD				
Param.		M0	M1a	M2a	M3a	M4a
			Fixed effects			
1	β_0	0.1714	2.795	1.667	1.509	1.509
	(a)	(0.1130)	(0.2454)	(0.2935)	(0.3165)	(0.3191)
K	β_K	3.541×10^{-5}	1.562×10^{-5}	2.572×10^{-5}	2.796×10^{-5}	2.796×10^{-5}
	(a)	(6.605×10^{-7})	(2.435×10^{-6})	(2.785×10^{-6})	(3.010^{-6})	(3.010×10^{-6})
	β_{K^2}	–	5.151×10^{-11}	2.906×10^{-11}	2.242×10^{-11}	2.242×10^{-11}
	(a)		(5.997×10^{-12})	(6.941×10^{-12})	(7.629×10^{-12})	(7.629×10^{-12})
W	β_{motor}	–	-0.1214	-0.1219	-0.1264	-0.1264
	(a)		(0.1076)	(0.1031)	(0.1025)	(0.1025)
	$\beta_{trailer}$	–	-1.710	-1.692	-0.1697	-0.1697
	(a)		(0.1046)	(0.1005)	(0.09997)	(0.1152)
	β_{leader}	–	0 (b)	0 (b)	0 (b)	0 (b)
H	β_{C2}	–	-0.6792	-0.2443	-0.4862	-0.4862
	(a)		(0.1473)	(0.3251)	(0.3288)	(0.3288)
	β_{C1}	–	0 (b)	0 (b)	0 (b)	0 (b)
			Random effects			
M	$\sqrt{d_M}$	–	–	0.6859	0.6838	0.6838
U	$\sqrt{d_U}$	–	–	–	0.2903	0.2903
V	$\sqrt{d_V}$	–	–	–	–	0.04042
			Scale			
	σ	2.086	1.896	1.815	1.803	1.803
			–2 restricted log likelihood			
		–	–	13414.6	13394.3	13394.3
			Akaike information criterium			
		14126.9	13503.6	–	–	–
			Number of parameters			
		3	7	8	9	10

(a) Approximate standard errors for fixed effects.

(b) This parameter is redundant.

Table 5: Restricted maximum likelihood estimates for the parameters of models M0 – M4b for the dependent variable change in the flange height (ΔFh).

		ΔF_h				
Param.		M0	M1b	M2b	M3b	M4b
			Fixed effects			
1	β_0	1.597	0.2386	0.004362	-0.01876	-0.01876
	(a)	(0.04645)	(0.1278)	(0.1436)	(0.1617)	(0.5048)
K	β_K	6.342×10^{-6}	1.881×10^{-5}	2.534×10^{-5}	2.608×10^{-5}	2.608×10^{-5}
	(a)	(2.715×10^{-7})	(2.348×10^{-6})	(2.666×10^{-6})	(3.122×10^{-6})	(3.122×10^{-6})
	β_{K^2}	–	2.680×10^{-11}	-2.250×10^{-11}	-2.664×10^{-11}	-2.664×10^{-11}
	(a)		(1.440×10^{-11})	(1.731×10^{-11})	(2.043×10^{-11})	(2.043×10^{-11})
	β_{K^3}	–	-1.748×10^{-16}	-7.748×10^{-17}	-7.269×10^{-17}	-7.269×10^{-17}
	(a)		(2.628×10^{-17})	(3.234×10^{-17})	(3.813×10^{-17})	(3.813×10^{-17})
W	β_{motor}	–	-0.09266	-0.09339	-0.09374	-0.09374
	(a)		(0.03929)	(0.03713)	(0.03692)	(0.03692)
	$\beta_{trailer}$	–	-0.7379	-0.7545	-0.7521	-0.7521
	(a)		(0.03818)	(0.03621)	(0.03604)	(0.6772)
	β_{leader}	–	0 (b)	0 (b)	0 (b)	0 (b)
H	β_{C2}	–	-0.1409	-0.3868	-0.3764	-0.3764
	(a)		0.05379	0.1615	0.1675	0.1675
	β_{C1}	–	0 (b)	0 (b)	0 (b)	0 (b)
			Random effects			
M	$\sqrt{d_M}$	–	–	0.3681	0.3804	0.3804
U	$\sqrt{d_U}$	–	–	–	0.1008	0.1008
V	$\sqrt{d_V}$	–	–	–	–	0.4782
			Scale			
	σ	0.8573	0.6920	0.6532	0.6495	0.6495
			–2 restricted log likelihood			
		–	–	6844.14	6830.07	6830.07
			Akaike information criterium			
		8297.42	6897.38	–	–	–
			Number of parameters			
		3	8	9	10	11

(a) Approximate standard errors for fixed effects.

(b) This parameter is redundant.

iii. The change in the flange thickness – ΔFt

The third dependent variable being modelled is the change in the flange thickness due to wear (ΔFt). As explained before in Figure 32, there is a lot of unexplained variability around the third-order polynomial describing the evolution of the diameter thickness due to wear with the kilometres since last turning/renewal. This variability is then explored again through LLMs, comparing the different specifications in Table 3 for the models M0–M4b. Table 6 provides the REML estimates for the parameters of the models explored. Note that, all the coefficients are statistically significant at the 5% significance level for all fixed effects. Comparing the variances with the total variance (σ2 + dM + dU + dV = 3.493), it is possible to find out that the measurement noise still captures 39.0%, the factor month of measurement (M) captures 55.4%, the factor unit (U) captures 1.4% and finally the factor vehicle (V) captures 4.2% of the total variance.

iv. The change in the flange slope – ΔqR

Finally, the fourth dependent variable being modelled is the change in the flange slope due to wear (ΔqR). As explained before in Figure 33, there is a lot of unexplained variability around the third-order polynomial describing the evolution of the diameter thickness due to wear with the kilometres since last turning/renewal. This variability is then explored again through LLMs, comparing the different specifications in Table 3 for the models M0–M4b. Table 7 provides the REML estimates for the parameters of the models explored. Note that, all the coefficients are statistically significant at the 5% significance level for all fixed effects. Comparing the variances with the total variance (σ2 + dM + dU + dV = 3.278), it is possible to find out that the measurement noise still captures 28.4%, the factor month of measurement (M) captures 50.0%, the factor unit (U) captures 0.1% and finally the factor vehicle (V) captures 21.5% of the total variance.

Table 6: Restricted maximum likelihood estimates for the parameters of models M0 – M4b for the dependent variable change in the flange thickness (ΔF_t).

		ΔF_t				
Param.		M0	M1b	M2b	M3b	M4b
			Fixed effects			
1	β_0	-0.08539	-3.805	-1.269	-1.670	-1.670
	(a)	(0.1049)	(0.3208)	(0.3071)	(0.3461)	(0.5168)
K	β_K	4.182×10^{-6}	4.803×10^{-5}	2.441×10^{-5}	3.096×10^{-5}	3.096×10^{-5}
	(a)	(6.131×10^{-7})	(5.894×10^{-6})	(4.955×10^{-6})	(6.114×10^{-6})	(6.114×10^{-6})
	β_{K^2}	–	-1.233×10^{-10}	-8.984×10^{-11}	-1.232×10^{-10}	-1.232×10^{-10}
	(a)		(3.614×10^{-11})	(3.226×10^{-11})	(4.016×10^{-11})	(4.016×10^{-11})
	β_{K^3}	–	2.889×10^{-17}	9.630×10^{-17}	1.506×10^{-16}	1.506×10^{-16}
	(a)		(6.596×10^{-17})	(6.023×10^{-17})	(7.482×10^{-17})	(7.482×10^{-17})
W	β_{motor}	–	0.1658	0.2093	0.2045	0.2045
	(a)		(0.09863)	(0.06685)	(0.06642)	(0.06642)
	$\beta_{trailer}$	–	0.2797	0.3028	0.2894	0.2894
	(a)		(0.09586)	(0.06521)	(0.06487)	(0.5465)
	β_{leader}	–	0 (b)	0 (b)	0 (b)	0 (b)
H	β_{C2}	–	-2.685	-3.342	-3.441	-3.441
	(a)		0.1350	0.5675	0.5722	0.5722
	β_{C1}	–	0 (b)	0 (b)	0 (b)	0 (b)
			Random effects			
M	$\sqrt{d_M}$	–	–	1.383	1.391	1.391
U	$\sqrt{d_U}$	–	–	–	0.218	0.218
V	$\sqrt{d_V}$	–	–	–	–	0.383
			Scale			
	σ	1.936	1.737	1.176	1.168	1.168
			–2 restricted log likelihood			
		–	–	10775.9	10776.4	10756.4
			Akaike information criterium			
		13638.6	12931.7	–	–	–
			Number of parameters			
		3	8	9	10	11

(a) Approximate standard errors for fixed effects.

(b) This parameter is redundant.

Table 7: Restricted maximum likelihood estimates for the parameters of models M0 – M4b for the dependent variable change in the flange slope (ΔqR).

		ΔqR				
Param.		M0	M1b	M2b	M3b	M4b
			Fixed effects			
1	β_0	-0.2254	-3.583	-0.5998	-0.6364	-0.6364
	(a)	(0.09199)	(0.2893)	(0.2647)	(0.2707)	(0.8819)
K	β_K	5.627×10^{-6}	3.729×10^{-5}	5.419×10^{-6}	5.836×10^{-6}	5.836×10^{-6}
	(a)	($5.376e \times 10^{-7}$)	(5.315×10^{-6})	(4.085×10^{-6})	(4.286×10^{-6})	(4.286×10^{-6})
	β_{K^2}	–	-3.994×10^{-11}	4.218×10^{-11}	4.084×10^{-11}	4.084×10^{-11}
	(a)		(3.259×10^{-11})	(2.659×10^{-11})	(2.800×10^{-11})	(2.800×10^{-11})
	β_{K^3}	–	-1.257×10^{-16}	-1.503×10^{-16}	-1.490×10^{-16}	-1.490×10^{-16}
	(a)		(5.949×10^{-17})	(4.964×10^{-17})	(5.227×10^{-17})	(5.227×10^{-17})
W	β_{motor}	–	0.08834	0.1316	0.1312	0.1312
	(a)		(0.08894)	(0.05494)	(0.05490)	(0.05490)
	$\beta_{trailer}$	–	0.3335	0.3558	0.3544	0.3544
	(a)		(0.08645)	(0.05360)	(0.05358)	(0.05358)
	β_{leader}	–	0 (b)	0 (b)	0 (b)	0 (b)
H	β_{C2}	–	-0.09933	-1.632	-1.634	-1.634
	(a)		(0.1218)	(0.5253)	(0.5265)	(0.5265)
	β_{C1}	–	0 (b)	0 (b)	0 (b)	0 (b)
			Random effects			
M	$\sqrt{d_M}$	–	–	1.288	1.280	1.280
U	$\sqrt{d_U}$	–	–	–	0.05599	0.05599
V	$\sqrt{d_V}$	–	–	–	–	0.8393
			Scale			
	σ	1.698	1.567	0.9661	0.9654	0.9654
			–2 restricted log likelihood			
		–	–	9508.04	9507.55	9507.55
			Akaike information criterium			
		12776.9	12254.3	–	–	–
			Number of parameters			
		3	8	9	10	11

(a) Approximate standard errors for fixed effects.

(b) This parameter is redundant.

This section provided a more comprehensive understanding on the topic of exploring wear trajectories of railway wheelsets. It introduced a new important variable - the flange slope (qR) - on the analysis of the wheelset degradation process which has a significant influence on the level of material removal during re-profiling. It also introduced the wheel hardness (H) as an explaining variable for the wheelset wear trajectories.

From the data analysis, the statistical patterns found were consistent with other train fleets and, therefore, it validated the statement that these "models can be applied to any fleet of vehicles".
The kilometres since last turning/renewal (K) is the variable with more influence in the wheelset wear trajectories among the variables analysed, but also the variable wheelset type (W) and wheelset hardness (H) are statistically significant. The factor month of measurement (M) exhibit a high variance in every model, which is likely to be due to adhesion variations (i.e. lower in Autumn), and it is something that goes in line with a previous study of Andrade and Stow (2016). Comparing to this previous research study for a different train fleet - where random effects associated with the factor month of measurement (M) exhibit a higher variance, followed by the factors unit (U) and vehicle (V) – the influence order of these same random effects was not the same in this case study, existing some variations in this order for each dependent variable here analysed. This could say that the factors that exhibit more variance to the wheelsets degradation trajectories depend on the fleets analysed, their technical specifications, as well as the climate conditions of each country.

As further steps for this specific case study analysis, two additional factors of the technician's influence and the wheelset damaged trajectories could be included. The assessment of data on the wear and damage trajectories can also be monitored by more sophisticated methods, such as doing some survival analysis for the data analysed. Moreover, the assessment of the wheel deterioration trajectories should consider the influence of the rail contact points and its deterioration processes as well (Lewis and Olofsson 2004). The inclusion in the models of the rail line data where this railway company operates would definitively improve this research study.

- **Statistical analysis of wheel wear of TfL London Underground dataset**

This final subsection provides a statistical analysis of wheelset condition data for the Jubilee Line 1996 Tube Stock (96TS) fleet from TfL London Underground (LUL), namely on the evolution of wear through: i) the diameter loss due to wear since last turning/renewal (Figure 34), ii) change in flange height due to wear (Figure 35), iii) change in flange thickness due to wear (Figure 36) and iv) change in flange slope due to wear (Figure 37).

LUL consists of a number of different lines, each with their own dedicated fleet of rolling stock, which have historically managed their wheelsets in different ways. The majority of fleets operated a 'run to fail' policy through the use of maintenance and inspection gauges. These gauges are used to identify when wheels are approaching flange height, thickness and hollowing limits, but if they still pass the inspection gauge, then they could be returned to service with wheel turning planned in before the next inspection. Even though the 'run to fail' policy was suitable when operating a limited service with spare train capacity, increases in timetables means that more trains are

required to operate the enhanced service, and hence, unscheduled maintenance must be avoided. A planned preventative wheel re-profiling regime
was required to ensure that a sufficient number of trains are available for service, which complies with the relevant standards. This also allows for wheel lathe demand to be accurately managed to ensure that demand does not exceed capacity.

In order to achieve this, a wheel profile monitoring programme was introduced to establish the principal mode of failure and the rate of degradation across each individual fleet so that a preventative wheel turning interval could be implemented within each train maintenance regime. The data presented in Figures 34 to 37 illustrates the outputs from this monitoring programme for the 96TS fleet.

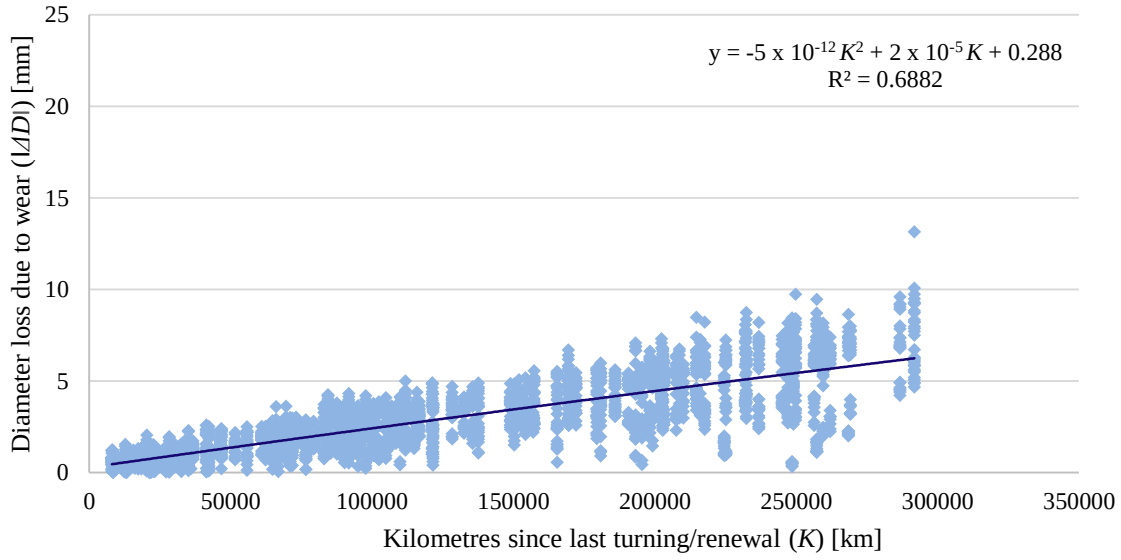


Figure 34: Diameter loss due to wear with the kilometres since turning/renewal [London Underground].

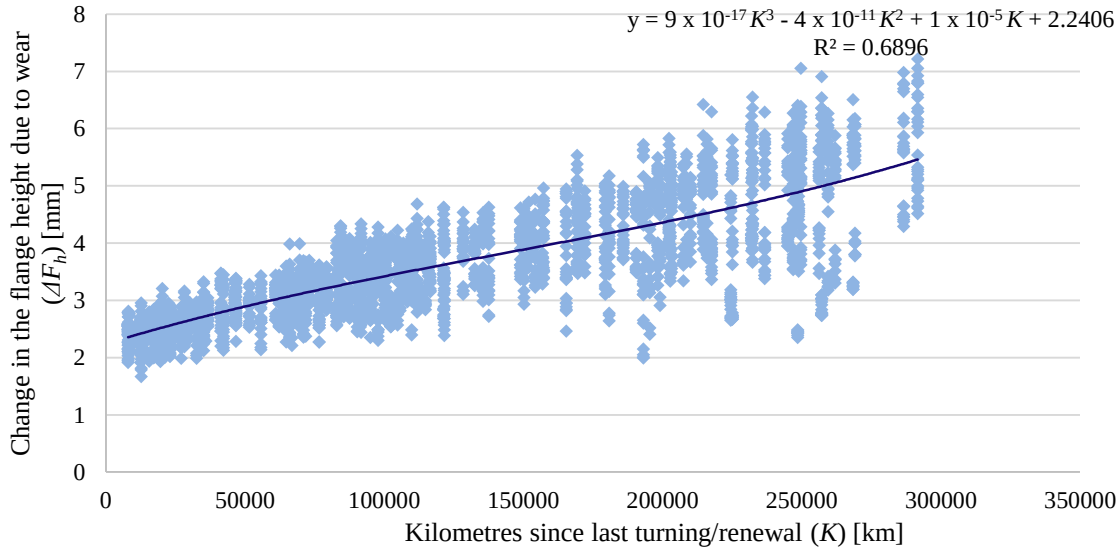


Figure 35: Change in the flange height due to wear with the kilometres since turning/renewal [London Underground].

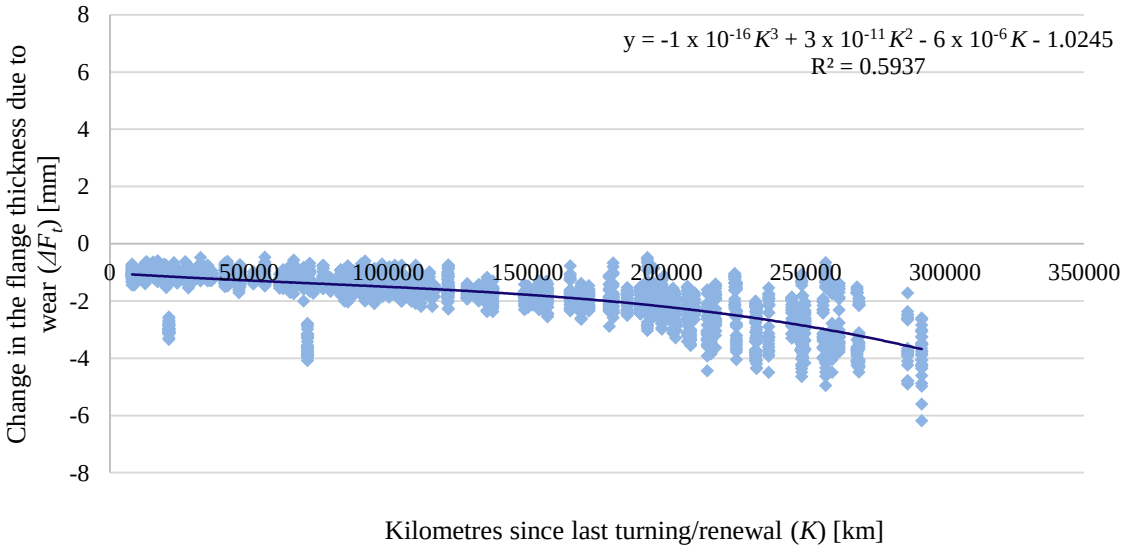


Figure 36: Change in the flange thickness due to wear with the kilometres since turning/renewal [London Underground].

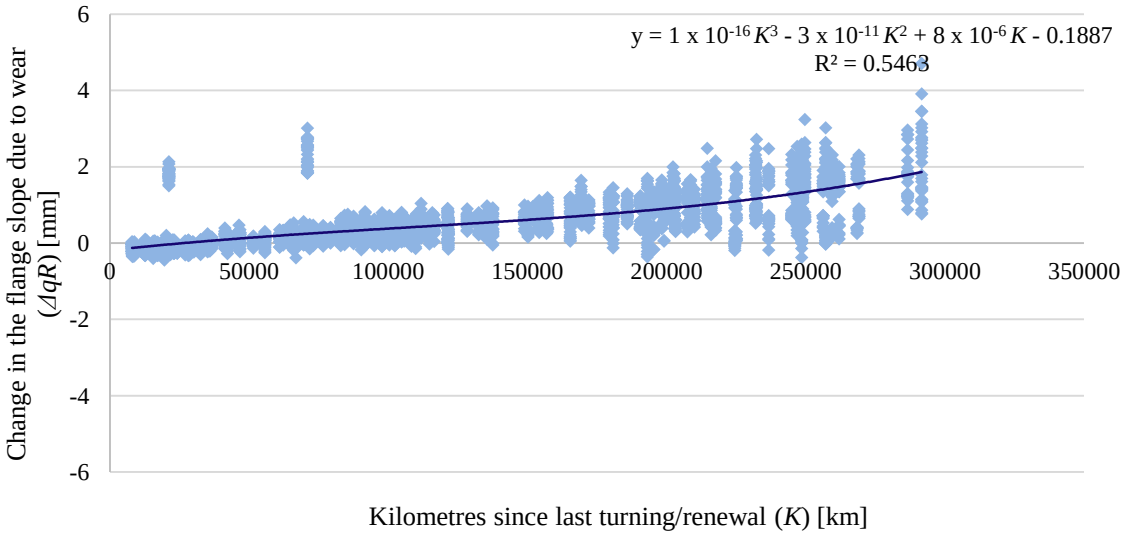


Figure 37: Change in the flange slope due to wear with the kilometres since turning/renewal [London Underground].

A comparison of wear data between the Fertagus case study (Figures 30-33) and LUL data (Figures 34-37) reveals that the LUL data has less variability around the trend lines and similar wear rates across all axles. It can be seen from the LUL data that the 96TS have higher rates of tread than flange wear (where the change in flange thickness (Figure 36) can actually be seen to grow since flange thickness is measured relative to a tread datum which moves down the angle face of the flange with tread wear which can appear as a growth in flange thickness), resulting in wheels primarily failing through a combination of wheel hollowing and zero (or negative) conicity (e.g. wheelset is no longer capable of generating any steering force).

Previously the fleet was re-profiled after every 570 days (≈ 256k km) which resulted in a number of wheels with levels of hollow tread wear and conicity which breached the criteria defined in the wheel standard. This increased the demand at the wheel lathe and resulted in large cut depths. Following a detailed review of wheel wear data and re-profiling strategy, it was recommended that the re-profiling interval was reduced to a frequency of 400 - 450 service days (≈ 180k – 200k km) to maximise the interval between turns and utilise the greatest amount of the wheel diameter possible.

It can be seen, based on the comparison of the Fertagus and LUL data, that although the LUL data provides less variation in the measured condition data significant work has already been undertaken to define an optimal planned preventative wheel re-profiling regime. The Fertagus case study was therefore selected for further assessment as it provides the greatest potential benefit from the application of the proposed decision support tools to support the derivation of an optimal wheelset maintenance strategy.

3.2 SURVIVAL MODELLING OF WHEELSET DAMAGE

A statistical survival analysis handles life time of individuals since their entrance in the study until any designated experience of interest (event/failure) occurs. It considers censored data, useful when the interest event does not occur in the observation time, i.e., when the wheelset is taken out of service without any damage for preventive maintenance. This section provides preliminary explanations for a better understanding on a damage occurrence survival analysis for railway wheelsets. Hereinafter in this section, the mathematical theories and definitions are in line with Kleinbaum and Klein (2012).

- **Survival curves:**

Considering a total live time of an individual T a continuous random variable with probability density function f(t), the survival probability S(t) of an individual surviving further than a time t ≥ 0 is given by:

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt$$

By definition, it is certain that every individual will survive in its initial live time, thus, $S(t = 0) = 1$. It is also natural that whenever an individual approaches its final life time it will get closer to a null survival probability, thus, $S(t \to +\infty) = 0$.

Therefore, a theoretical survival curve would be similar to the one represented in Figure 38. Using statistical lifetime data for estimating survival curves, a non-parametric statistical method relies on a product limit estimator called Kaplan-Meier estimator. This method uses statistical real data for calculating survival probabilities and, thus, deals with real-life problems as censored data.

The Kaplan-Meier estimator computes the probability of an individual surviving at some determined time (t_i) knowing that the individual has survived all the previous time intervals. The probability within a time interval is considered the same and its values are determined by the data available within each interval. Kaplan-Meier Estimator turns into a product of probabilities, since it considers the independence of events between time intervals:

$$\hat{S}(t_i) = \prod_k^i \hat{P}(T > t_k | (T \geq t_k)$$

In which: $i, k \in \{1,2,3, \dots\}$.

Therefore, the Kaplan-Meier survival curves are step functions obtained from the practical survival probabilities derived from the Kaplan-Meier estimators (Figure 38 (b)). They are finite curves

depending on the available data last lifetime-length (study end), thus, they usually do not go all the way down to zero.

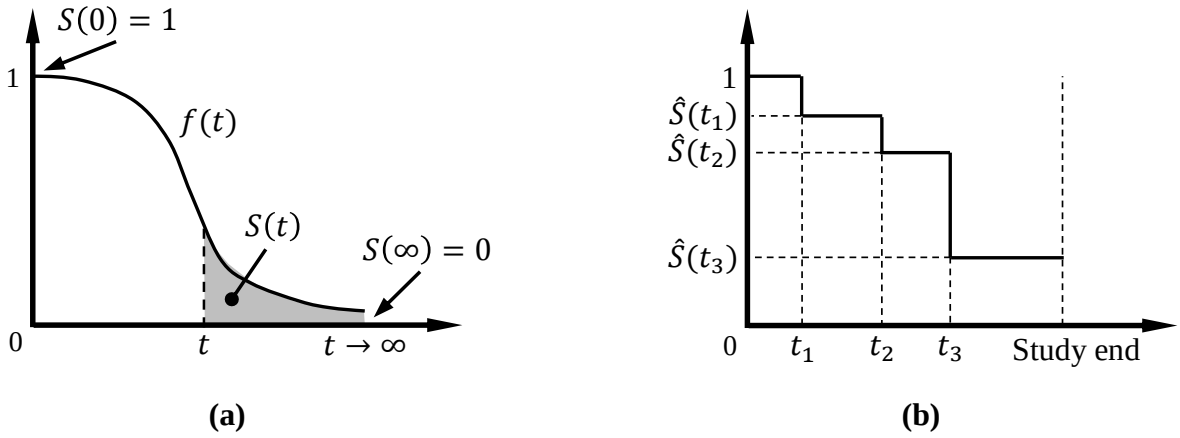


Figure 38: Survival curves: (a) Theoretical survival curve; (b) Kaplan-Meier survival curve

In this survival analysis for railway wheelsets, the event studied is the damage occurrence (it is only considered wheel damages, since the occurrence of preventive maintenance actions due to axle damages is minimal); the time variable is the number of kilometres since each wheelset last turning or renewal (kst); the censoring event is a non-damage turning activity. Figure 39 illustrates a follow-up time for six representative wheelsets. Note that, using this type of modelling, every individual starts the observation time and every censored data is right-censored.

There is a kst interval around an established kst value determined by each train operating company, where preventive turnings occur. After this last interval, all life records of the possible surviving wheelsets are censored by the turning activities.

Figure 39: Railway wheelsets follow-up time

Two databases are used in this survival analysis:

i. The first database comes from wheelset turning maintenance operations, of a fleet of 18 EMU trains of a single type or class, between October 2000 up to June 2015 (i.e. a 16-year interval). Each unit has four vehicles and each vehicle has eight wheels (i.e. four wheelsets). This database is from a Portuguese passenger train operating company and took several visits to their maintenance yards to compile it.

ii. The second database is from a British passenger train operating company, having wheel data from December 2006 up to July 2012 (i.e. a 7-year interval) and containing information of 51 EMU trains of a single type or class. Each unit has three vehicles and each vehicle has eight wheels (i.e. four wheelsets). This database was previously discussed in Andrade and Stow (2016, 2017a, 2017b), without looking at a survival modelling approach.

The Portuguese train operating company does not distinguish between a case of preventive maintenance operation or a case of corrective maintenance operation (the case of correcting a damage) when doing turning operations, neither registers in which specific wheelset occurs the damage. These are all limitations for this survival modelling.

To not include in this analysis survival probabilities mixing damage occurrences with preventive turnings, it is only considered the data until 100,000 kst in the estimation of the Kaplan-Meier survival curve (black coloured in Figure 40) for the Portuguese train operating company, since this is the time interval that is known for sure to be out of the preventive turning zone represented in Figure 40. However, when this company has to correct wheel damage, they have to turn the whole vehicle (because wheel diameters cannot vary much from the opposite side wheel, cannot vary much between wheels of the same bogie and wheelsets between bogies). This Kaplan-Meier curve is considering every wheel in this situation as a damaged wheel, introducing some bias to the curve. These considered survival probabilities were then adjusted to a Weibull distribution (blue coloured in Figure 40), a very flexible distribution, widely used in survival analysis since it well represents theoretical survival curves. Its parameters, as well as some likelihood criteria can be seen in Table 8.

Table 8: Portuguese train operating company Weibull survival curve parameters and some likelihood criteria

Distribution	AIC	BIC	Log Likelihood	Parameter 1	Parameter 2
Weibull	3556	3568	-1776	1.946	395.984

Due to these previous database limitations, a survival analysis on a different train operating company was performed with the intention to have a comparison reference for the wheelset damage trajectories using a more reliable wheelset damage database.

This British train operating company database is more reliable in a way that it does not have the previous presented limitations and it has competing wheel damage risks.

When using railway wheelset real operating data, every damage registered in the preventive turning kst interval (blue zone in Figure 39) will be overweighting the calculus of the survival probabilities using the Kaplan-Meier method, due to the large amount of preventive turning censored data in those intervals (please see Kleinbaum and Klein (2012) for better understanding of the Kaplen Meier estimator survival probabilities calculus). In a Kaplan-Meier curve, this is represented in large time-steps as the ones observed in the Kaplan-Meier curve for the British train operating company after around 250,000 kst (grey coloured in Figure 40). Therefore, it was performed an iterative approach for choosing a maximum kst limit for analysing the data without falling into the preventive turning zone. Several models were derived and it was chosen the model with the best trade-off between a likelihood criteria and a good theoretical survival curve shape. Some of these models can be seen in Table 9 and the model chosen was the Weibull distribution using as maximum kst 250 thousand km. This model is depicted in orange colour in Figure 40.

Table 9: British train operating company survival curves parameters and some likelihood criteria.

Scenario	Distribution	AIC	BIC	Log Likelihood	Parameter 1	Parameter 2
250	Weibull	27125	27139	-13661	1.357	437.598
	Normal	27299	27313	-13648	280.610	135.030
	Log-Normal	27480	27493	-13738	6.099	1.451
275	Weibull	27215	27228	-13605	1.331	453.070
	Normal	27443	27456	-13719	289.032	141.052
	Log-Normal	27554	27568	-13775	6.126	1.467
300	Weibull	27257	27271	-13627	1.316	461.501
	Normal	27524	27538	-13760	293.501	144.527
	Log-Normal	27586	27599	-13791	6.137	1.474
325	Weibull	27274	27288	-13635	1.310	465.046
	Normal	27560	27573	-13778	295.367	146.099
	Log-Normal	27598	27612	-13797	6.142	1.477
350	Weibull	27279	27292	-13637	1.308	465.950
	Normal	27570	27584	-13783	295.973	146.631
	Log-Normal	27602	27615	-13799	6.143	1.478
375	Weibull	27280	27293	-13638	1.308	466.094
	Normal	27573	27587	-13785	296.219	146.789
	Log-Normal	27602	27616	-13799	6.143	1.478
400	Weibull	27280	27294	-13638	1.308	466.240
	Normal	27574	27588	-13785	296.248	146.813
	Log-Normal	27602	27616	-13799	6.143	1.478

Observing the grey functions in Figure 40 it is possible to see that the Kaplan-Meier curve does not look like a theoretical survival curve and looks like a composition of competing damage risks. Grouping the risks as in Figure 41, it is possible to see that the Weibull distributions fit well in the survival curves. If doing a serial Weibull function, it is possible to see that it is obtained a much better survival curve, which seem to have a similar behaviour has the one derived for the Portuguese train operating company case.

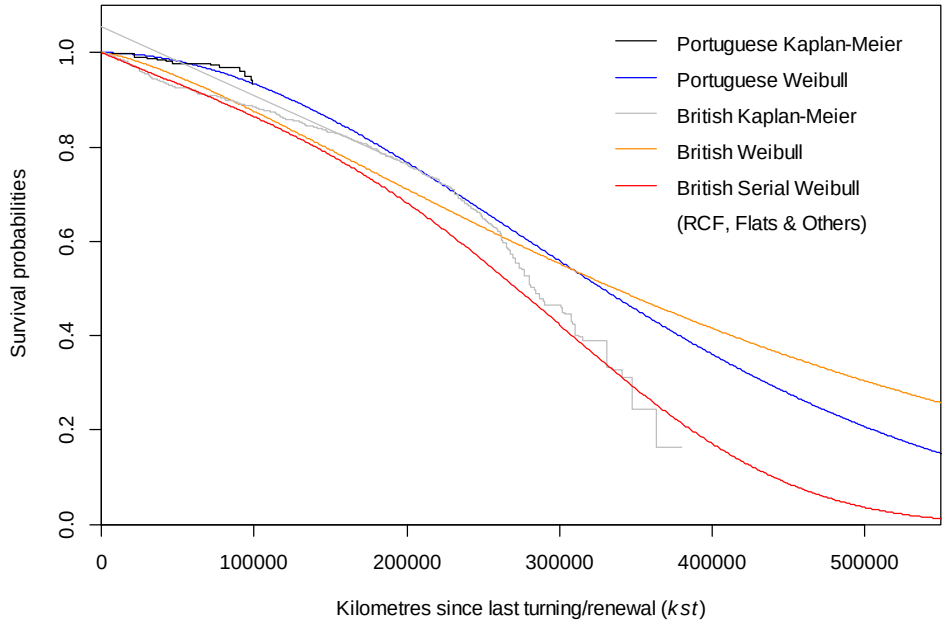


Figure 40: Kaplan-Meier and Weibull survival curves

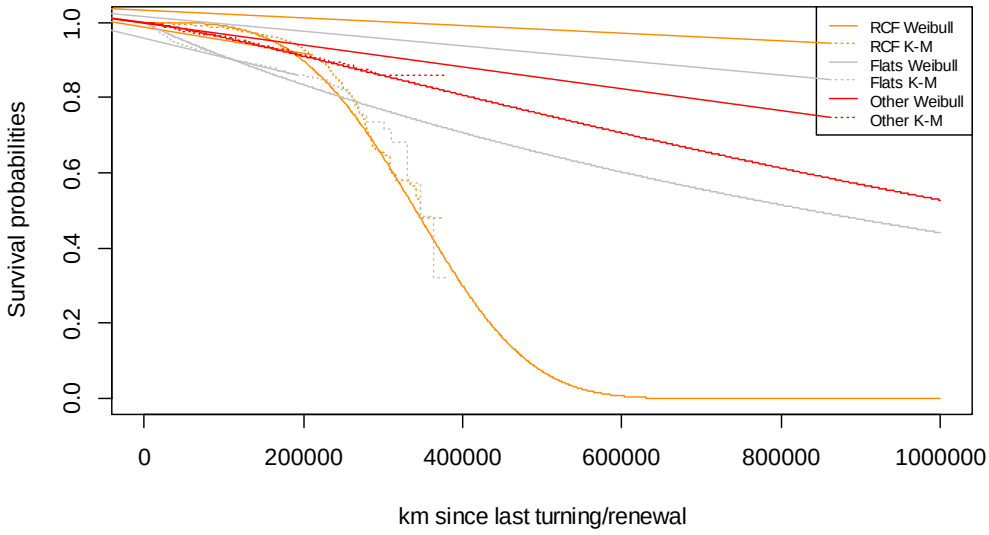


Figure 41: British train operating company competing damage risks survival curves

Figures 40 and 41 can then be the basis to estimate the transition probability matrixes, namely the associated hazard rates to damage occurrence. However, in the next section 3.3, the hazard function will be estimated directly (using the Cox Proportional Hazards Model (CPHM) and estimate Figure 46).

3.3 MARKOV DECISION PROCESS APPROACH

A Markov decision process (MDP) approach is applied to railway wheelset maintenance, using data of wheelset maintenance activities from a Portuguese train operating company.

- **Model formulation**

This MDP modelling considers:

(i) The wheel diameter (*D*) which is a key indicator of the lifecycle stage that a given wheel is at a certain epoch (*n*);
(ii) The wheel damage occurrence (such as rolling contact fatigue (RCF), flats or cavities) which is responsible for the most severe maintenance actions in railway wheelsets, shorten significantly their lifecycles;
(iii) The kilometres since last turning/renewal (*kst*) operation of each wheelset;
(iv) Three possible maintenance actions (*a* = 1, 2, 3):
  - 'Do nothing' (*a* = 1): the wheelset is ok and it goes back to service in the same state;
  - 'Renewal' (*a* = 2): the corrective or preventive maintenance actions would need to go beyond the scrap diameter, and so the wheel must be replaced by a new one;
  - 'Turning' (*a* = 3): the wheelset goes to a turning lathe for its shape being replaced to values within the standards and it suffers a reduction/loss in its diameter.

The final objective of the present model is to determine an optimal wheelset maintenance strategy based on wheel deterioration processes in an MDP framework. The maintenance costs in the long run are minimized and it is provided a decision map depending on the wheel diameter, damage occurrence and kilometres since last turning/renewal. This decision map is expected to contain decisions of predictive maintenance which can be included in the train operating company wheelset reprofiling policy.

This MDP is derived over an infinite planning horizon and the MDP is considered stationary, i.e. (i) the transition probabilities are assumed to be constant over time, and thus, the Markov transition matrices (MTMs) are independent of the epoch at which the transition occurs; (ii) the policy is independent of time.

The following 'State space' subsection starts explaining the state space used in the modelling. Then, subsection 'Estimation of MTMs' explains the estimation of the MTMs for each possible action. Subsection 'Reward/cost function' discusses how the reward/cost functions are defined/estimated. The next subsection 'Optimal policy' provides the optimal maintenance policy,

which maps the best possible action depending on the condition state of the wheelset. Finally, the last subsection discusses the main conclusions withdrawn from the analysis of the decision map obtained.

- **State space**

The state space is defined based on three main chosen indicators for the wheelset states: (i) wheel diameter (D), (ii) the mileage since the last turning (kst) and (iii) the occurrence of damage. The wheel diameter is varying from an initial diameter ($D_{initial}$) of 920 mm until a scrap diameter (D_{scrap}) of 850 mm and the diameter categories are grouped in intervals with amplitudes of 1 mm (i.e. 70 different levels). The kilometres since last turning/renewal (kst) vary from 0 up to 350,000 km in intervals of 10,000 km, which is the considered MDP time step (t). (i.e., 36 different levels). Finally, a wheelset can be in a state of damage or not. Consequently, a total of 2590 different states, $s \in \{s_1, s_2, \ldots, s_{2590}\}$, is defined. Note that the 70 states with damage are kept at the end of the state space, but without the extension of each damaged state depending on the kilometres since last turning, as the transitions from damaged states to non-damaged states are compulsory, because once the damage is detected, it must be removed.

- **Estimation of MTMs**

An MTM has to be defined for each possible action. This section is divided in three subsections explaining the estimation of the 'Do nothing' MTM ($\boldsymbol{P}_1$), the 'Renewal' MTM ($\boldsymbol{P}_2$) and the 'Turning' MTM ($\boldsymbol{P}_3$). As explained earlier, this study considers the wheel diameter (D) as the main indicator of the wheel's lifecycle stage. In this analysis and as suggested in Figure , the wear in the wheelset - measured as the diameter loss due to wear (ΔD) - is assumed to be independent of the wheel initial diameter after a renewal or reprofiling. This assumption is reasonable as the hypothesis of independence cannot be rejected at a significance level of 0.05.

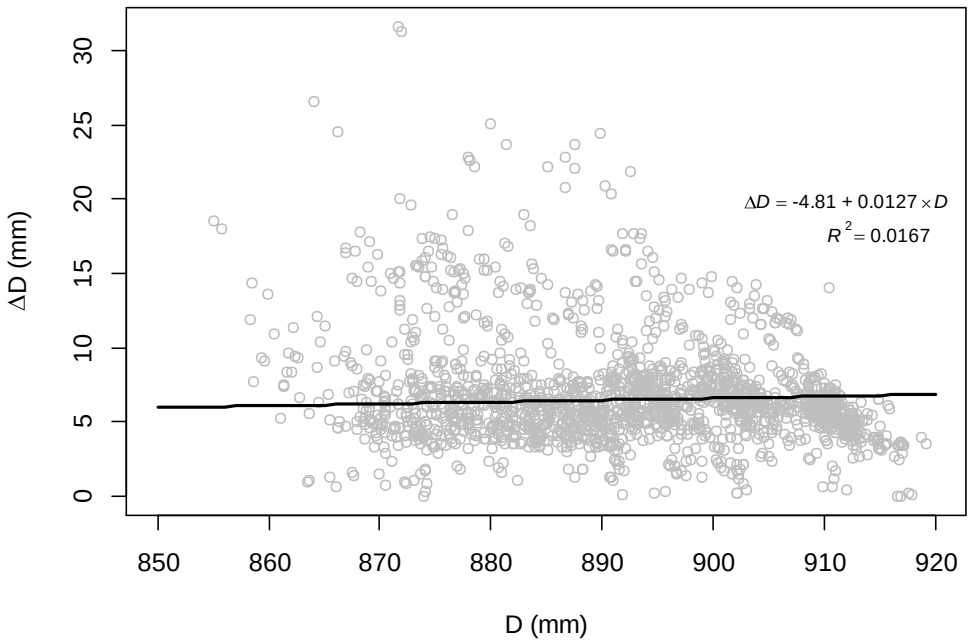


Figure 42: Diameter loss due to wear (ΔD) for different diameters (D).

Considering a homogenous Markov chain, the transition matrix is decomposable into several sub-transition matrixes – in a diagonal block form (Yin and Zhang, 2006). For a Markov chain with a finite but large state space, the decomposition of the transition matrix tends to follow the most attractive approach. Having said that, the underlying problem of estimating MTMs can be divided into sub-problems that can be solved independently. This kind of approach will be followed in the next subsections with the estimation of sub-transition matrices.

- *'Do nothing' action (a=1)*

The 'Do nothing' action considers that the only possible way to increase a wheel's diameter is through renewal. Furthermore, as data suggests, abrupt decreases in diameter (due to wear) are very unlikely to happen. Therefore, a simplification is considered where the only possible transitions for a given state (not considering damage states transitions) is to move to a state of diameter immediately below (with probability θ) or stay in the same state (with probability 1 - θ), see Figure 43 for a schematic representation.

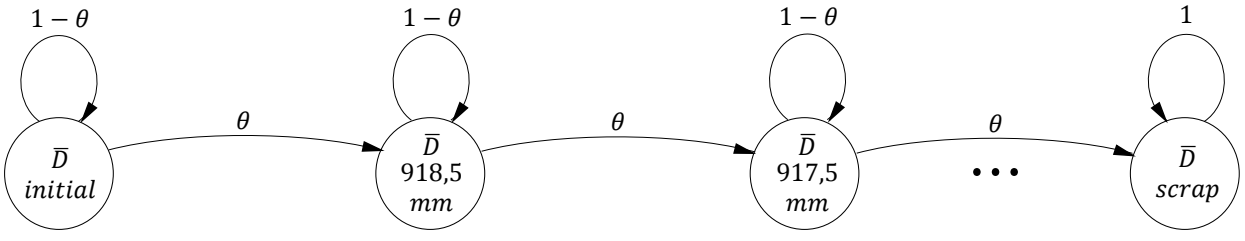


Figure 43: Transitions between states without damage depending on the parameter θ for the 'Do nothing' action, adapted from Braga and Andrade (2019).

Regarding the tread change diameter due to wear, it is possible to predict the mean value ($\overline{\Delta D}$) with the Markovian approach each time step of 10,000 km as

$$\overline{\Delta D}_{(n)} = \boldsymbol{X}_0 \boldsymbol{P}^n \boldsymbol{\Delta D}$$

To derive these scalar mean values, the following variations in the wheel diameter for the wear states were considered

$$\boldsymbol{\Delta D} = \begin{bmatrix} 0 & 1 & 2 & \cdots & 69 \end{bmatrix}^{\mathrm{T}} (mm)$$

These were the possible variations for the wheelset diameters derived from the diameter state representative values considered in the sample (i.e. diameter categories mean values from $\overline{D}_{initial}$ = 919.5 mm up to $\overline{D}_{scrap}$ = 850.5 mm).

The initial state of the wheelset is:

$$\boldsymbol{X}_0 = \begin{bmatrix} P(\Delta D = 0) & P(\Delta D = 1) & \cdots & P(\Delta D = 69) \end{bmatrix}$$

Algebraically, considering the chosen transition matrix (with zeros in all entries that do not belong to the diagonal or upper diagonal), it is possible to derive, for several values of θ in a grid from 0.05 up to 0.5 and considering the *n*-step transition probabilities (i.e., the probability that a process in state *i* will be in state *j* after *n* additional transition), several lines and see which one best fits the distribution. From a quick inspection, the one with θ = 0.35 seems to be providing the best approximation to the black dashed line. In fact, θ = 0.36 (dashed line) provides the closest fit to the regression line, according to the ordinary least squares for the slope of the simple linear regression without an intercept (Figure 44). Note that the assumption of not including an intercept (or in other words assuming that the intercept is equal to zero) is aligned with no wear, that is, ΔD = 0, when a wheelset is new or just turned and it has no kilometres since last turning/renewal. Finally, the value of θ = 0.36 is chosen.

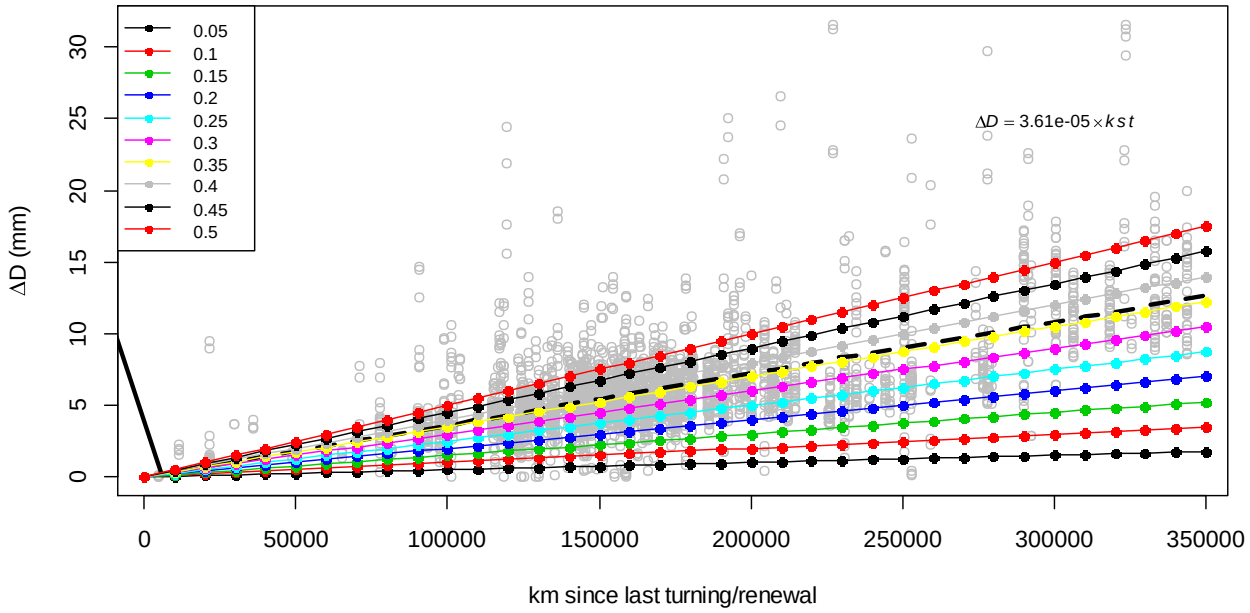


Figure 44: Diameter loss due to turning (ΔD) with kilometres since turning/renewal, applying Markovian approaches and linear regression without intercept

Next, transition probabilities to states with damage must also be derived. The main assumption is that a wheel stays damaged without a change in its diameter, since, in theory, once wheelset damage is detected the vehicle must be removed from service and the wheelset reprofiled. Therefore, transitions from wheels without damage to damaged states are schematically represented in Figure 45.

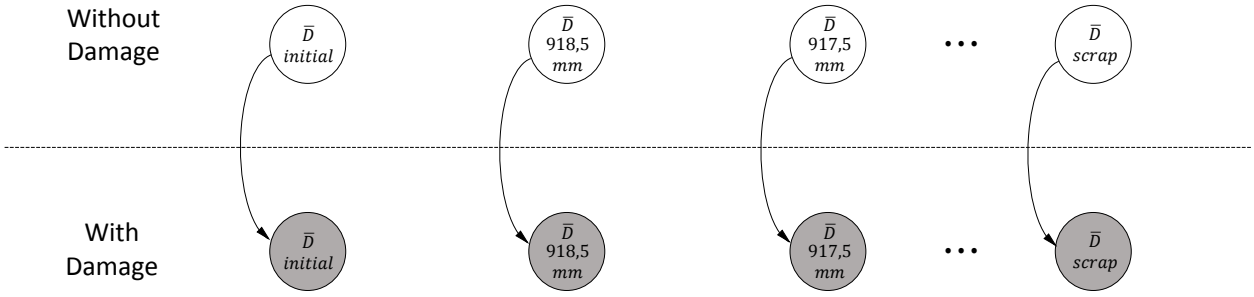


Figure 45: Considered transition probabilities to states with damage, adapted from Braga and Andrade (2019)

For deriving the damage probabilities, a Cox Proportional Hazards Model (CPHM) is implemented (Cox, 1972; Cox and Oakes, 1984). Due to the database limitation providing reliable information regarding the damage occurrence, another database from previous references (Andrade and Stow, 2016, 2017a, 2017b; Braga and Andrade, 2019), and from a different train operating company, is used for deriving the CPHM.

This model is a regression-type approach to survival curves, whenever the use of covariates (categorical or continuous) is needed. In this case study, this model is used for deriving the survival probabilities of a wheel given its diameter value.

The hazard function *h(kst, D)* in the CPHM for a wheelset (one observation) at a given value of $kst = k$ and *D* = *d* can be calculated as:

$$h(k,d) = h_0(k) \times e^{\beta(d)}$$

In the hazard equation above, the covariate is the tread diameter (*D*) and its coefficient β measures its size effect. The quantity e^{β} is the hazard ratio linked to the covariate *D*, and it was shown to be statistically significant, with the upper term of the 0.95 confidence interval being 0.982 (slightly below 1), indicating that as the tread Diameter (*D*) increases, the hazard decreases and, hence, length of survival increases, i.e. new wheelsets have longer survival than wheelsets whose diameter is close to the scrap diameter.

The cumulative hazard function *H(kst, D)* in the CPHM for a wheelset (one observation) at a given value of $kst = k$ and *D* = *d* can be calculated as:

$$H(k,d) = H_0(k) \times e^{\beta(d)}$$

In both equations above, $h_0(k)$ and $H_0(k)$ are baseline and cumulative baseline hazards, respectively, which are obtained when the value of the covariate *d* is set to 0 in the corresponding equations. Also, for a given wheel:

$$S(k,d) = (S_0(k))^{\exp\{\beta(d)\}}$$

The equation above shows how the Cox Model computes the survival probabilities based on a survival function $S_0(k)$ corresponding to the baseline hazard $h_0(k)$. The survival probability, for a fixed wheel diameter, at a given *kst* = *k* represents the probability of survival beyond *k*, i.e.:

$$S(k) = P(kst > k) \quad , \quad k > 0$$

Starting from the survival curves, it is more intuitive to understand the computation of the hazard rates. The hazard function at a fixed diameter, *h(kst)*, assesses the instantaneous risk of failure at *kst* = *k*, conditional on survival to that *kst* or in mathematical notation:

$$h(k) = \lim_{\Delta k \to 0} \frac{\Pr[(k \le kst < k + \Delta k)]}{\Delta k} = \frac{f(k)}{S(k)} = -\frac{S'(k)}{S(k)}$$

The equation above relates the hazard function to the survival function. For this work, hazard rates derived from the CPHM are displayed in Figure 46, where the line closer to the origin (with lowest

hazard rates) corresponds to the hazard function for the highest tread diameter representative value, $\overline{D}_{initial}$ of 920 mm, and the upper curve corresponds to the hazard function for the lowest tread diameter representative value, $\overline{D}_{scrap}$ of 850 mm.

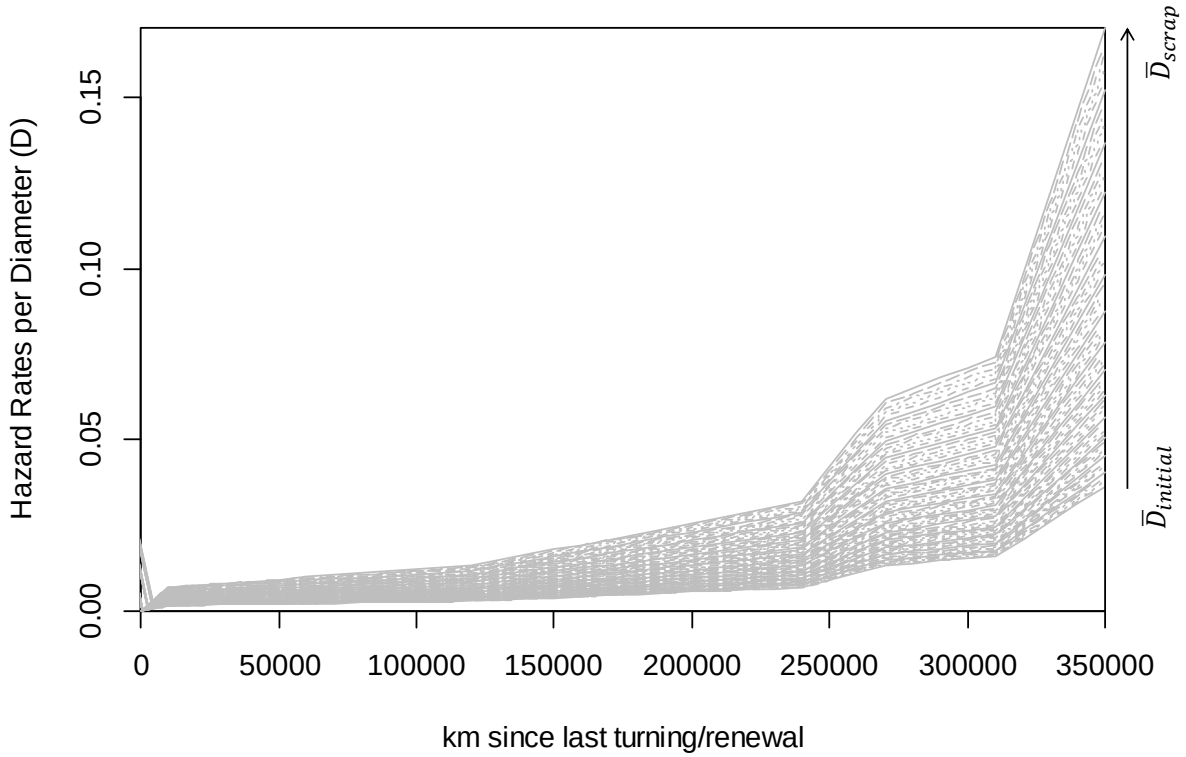


Figure 46: Estimated survival probabilities per diameter group representative values with kilometres since last turning.

The probability of occurring damage in a wheel of a given diameter at a certain *kst* is taken as simply the discretized values of hazard curves in Figure 46.

A sub-transition matrix for the damage probabilities, considering all 2520 states without damage to the 70 states with damage can be represented in the following way:

$$
\begin{array}{c}
\qquad\qquad \bar{D}_{\substack{initial\\damaged}} \qquad\qquad\qquad\qquad\qquad\qquad \bar{D}_{\substack{scrap\\damaged}} \\
\boldsymbol{P_D} = \left[p^{1}_{i,j} \right] = \left[\begin{array}{cccc}
p(damage) & 0 & \cdots & 0 \\
0 & p(damage) & \ddots & \vdots \\
\vdots & \ddots & p(damage) & 0 \\
0 & \cdots & 0 & p(damage) \\
\hline
p(damage) & 0 & \cdots & 0 \\
0 & p(damage) & \ddots & \vdots \\
\vdots & \ddots & p(damage) & 0 \\
0 & \cdots & 0 & p(damage) \\
\hline
p(damage) & 0 & \cdots & 0 \\
0 & p(damage) & \ddots & \vdots \\
\vdots & \ddots & p(damage) & 0 \\
0 & \cdots & 0 & p(damage) \\
\hline
p(damage) & 0 & \cdots & 0 \\
0 & p(damage) & \ddots & \vdots \\
\vdots & \ddots & p(damage) & 0 \\
0 & \cdots & 0 & p(damage)
\end{array} \right]
\begin{array}{l}
\left.\begin{array}{l} \bar{D}_{initial} \\ \\ \\ \bar{D}_{scrap} \end{array}\right\} kst\ 0\ km \\
\left.\begin{array}{l} \bar{D}_{initial} \\ \\ \\ \bar{D}_{scrap} \end{array}\right\} kst\ 10k\ km \\
\left.\begin{array}{l} \bar{D}_{initial} \\ \\ \\ \bar{D}_{scrap} \end{array}\right\} \vdots \\
\left.\begin{array}{l} \bar{D}_{initial} \\ \\ \\ \bar{D}_{scrap} \end{array}\right\} kst\ 350k\ km
\end{array}
\end{array}
$$

It has been well accepted that damage occurrence is considered independent of wear (ΔD), hence it follows that the joint probability of damage and wear is equal to the product of the marginal probabilities, as follows:

$$P(wear \cap damage) = P(wear) \cdot P(damage)$$

Therefore, the sub-transition matrix for wear needs to be modified so that probabilities of damage are incorporated. The non-zero entries of the matrix can be computed as:

$$
\begin{cases}
p_{i,i+70} = (1-\theta)\left(1-p(damage)\right) & ;i = j + 70k \\
p_{i,i+71} = \theta\left(1-p(damage)\right) & ;i = j + 70k \qquad ;\ j = 1,2,\ldots 69\ ; k = 0,1,\ldots,34 \\
p_{i,i+70} = 1 - p(damage) & ;i = 70(k+1) \\
p_{i,i} = 1 - p(damage) & ;i = 2521, 2522, \ldots, 2590
\end{cases}
$$

In a matrix form, the final MTM for the 'Do nothing' action ($\boldsymbol{P}_1$) is a 2590 by 2590 matrix composed by the sub-transition matrices in a diagonal form, as follows:

$$\boldsymbol{P}_1 = \left[p_{i,j}^{1} \right] = \begin{bmatrix} \overset{0}{\underset{km}{}} & \overset{10k}{\underset{km}{}} & \overset{20k}{\underset{km}{}} & \cdots & \overset{340k}{\underset{km}{}} & \overset{350k}{\underset{km}{}} & \overset{states}{\underset{with\ \ damage}{}} & \\ \mathbf{0} & \boldsymbol{P}_W^{(70\times70)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \overset{0}{\underset{km}{}} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{P}_W^{(70\times70)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \overset{10k}{\underset{km}{}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \boldsymbol{P}_D^{(2520\times70)} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{P}_W^{(70\times70)} & \mathbf{0} & & \overset{340k}{\underset{km}{}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{P}_W^{(70\times70)} & & \overset{350k}{\underset{km}{}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}^{(70\times70)} & \overset{states\ with\ dam.}{} \end{bmatrix}^{(2590\times2590)}$$

- *'Renewal' action (a=2)*

Concerning the 'Renewal' action, regardless of the current state of the wheel (damaged or undamaged), transitions to the initial state are assumed to be certain, as described in Figure 47.

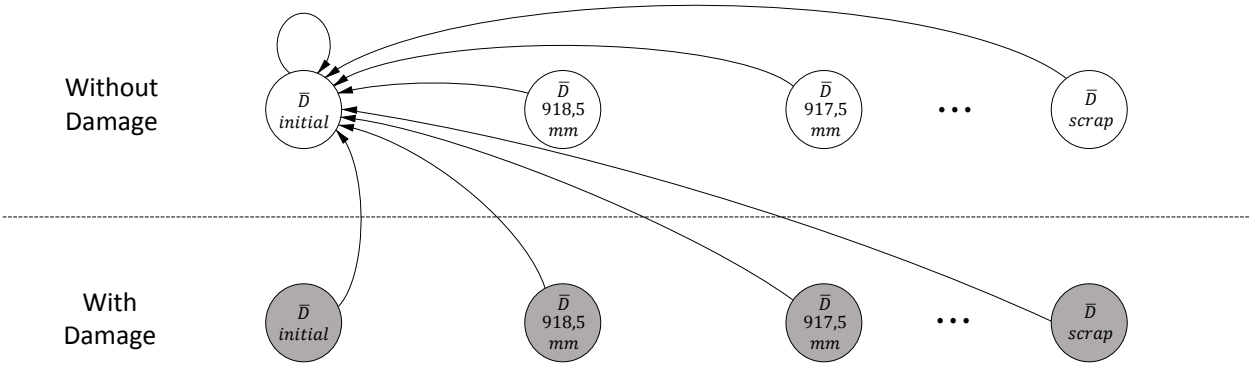


Figure 47: Transitions between states for the 'Renewal' action, adapted from Braga and Andrade (2019).

Therefore, the MTM for the 'Renewal' situation ($\boldsymbol{P}_2$) is a 2590 by 2590 matrix as follows:

$$\boldsymbol{P}_2 = \left[p_{i,j}^{2} \right] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}^{(2590\times2590)}$$

- *'Turning' action (a=3)*

The possible transitions between states for the 'Turning' action, in theory, are schematically represented in Figure 48.

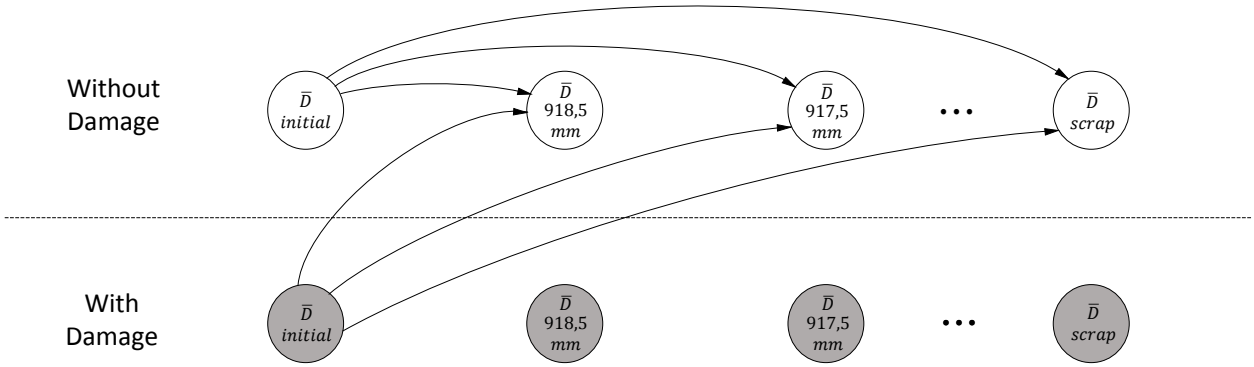


Figure 48: Transitions between states for the 'Turning' action, adapted from Braga and Andrade (2019).

When turning a wheelset, there is a distinct loss in the diameter due to turning (reprofiling of the wheel) if it is a situation of correcting damage (e.g. wear and surface fatigure damage) or if it is a situation of preventive turning. In the case of correcting damage, the diameter loss tends to be significantly larger on average and with a higher dispersion as depicted in Figure from Braga and Andrade (2019).

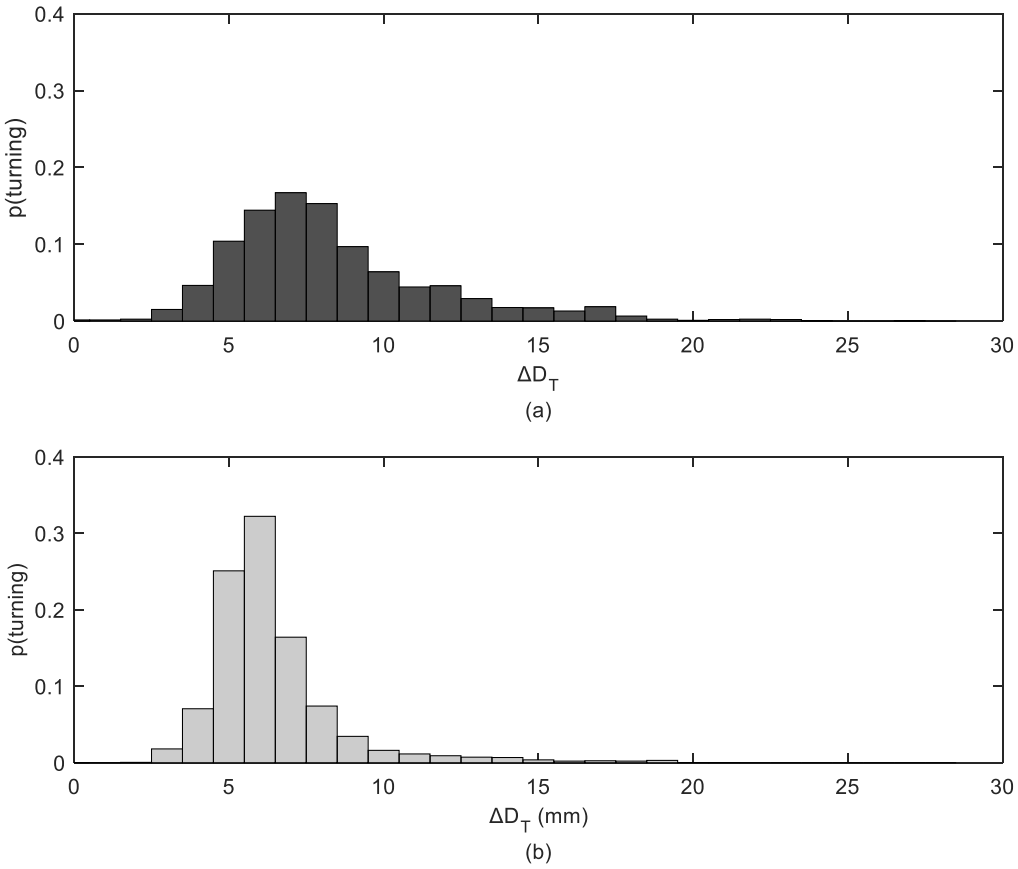


Figure 49: Histograms of the loss in diameter due to turning (ΔDT) in a wheelset: (a) with damage and (b) without damage (Braga and Andrade, 2019)

The Portuguese train operating company database does not differentiate between turning situations of correcting damage or not. Due to this database limitation, the diameter loss due to turning (ΔD_T) probability distributions for damaged and undamaged wheels were considered the same as in Braga and Andrade (2019).

The probabilities represented in Figure 49 withdrawn from Braga and Andrade (2019) were calculated using the relative frequency from past samples as an approximation of the transition probabilities, that is:

$$p(turning) = \frac{n_j}{N}$$

In which n_j is the number of wheelsets that transit to a class j of diameter loss and N is the total number of wheelsets.

The MTM for the 'turning action' assumes that the transitions to next states are limited, meaning that a transition from a state to another one with a great loss in the diameter does not happen at some point (according to Figure 49, it is defined 30 mm as the maximum loss in diameter possible). Therefore, regarding transitions from one state to another, the probabilities are composed by zeros

to states before the current one and zeros for states after the current one that the 'turning' action 'cannot reach'.

When a wheelset is turned, it goes back to a state where kilometres since last turning/renewal (*kst*) are zero, and if it has damage it goes to a state without damage, since once the damage is detected it must be removed. As it is not possible to turn a wheelset beyond the scrap diameter, when the wheelset is in a scrap diameter state, at some point of its kilometres since last turning (*kst*), and the histograms of Figure 49 (a) indicate diameter losses that go beyond the scrap diameter for that final state, the probabilities of the remaining transitions are summed up becoming the probability value for the wheelset to stay at the final state, that is, the scrap diameter. Having said that, it is possible to compose the sub-transition matrix for the 'turning' action from states without damage ($\boldsymbol{P}_{TND}$):

$$\boldsymbol{P}_{TND} = \left[p^{3}_{i,j} \right] = \begin{bmatrix}
p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & \cdots & p_{1,29}^{3^{TND}} + p_{1,30}^{3^{TND}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & \cdots & p_{1,28}^{3^{TND}} + p_{1,29}^{3^{TND}} + p_{1,30}^{3^{TND}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \sum_{j=1}^{30} p_{1,j}^{3^{TND}} = 1
\end{bmatrix}$$

In the same way, using now the probability values withdrawn from Figure 49 (b), it is possible to compose the sub-transition matrix for the 'turning' action from states with damage ($\boldsymbol{P}_{TD}$):

$$\boldsymbol{P}_{TD} = \left[p_{i,j}^{3} \right] = \begin{bmatrix}
p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & p_{1,3}^{3^{TD}} & \cdots & p_{1,30}^{3^{TD}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TD}} & p_{1,2}^{3^{TD}} & \cdots & p_{1,29}^{3^{TD}} + p_{1,30}^{3^{TD}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TD}} & \cdots & p_{1,28}^{3^{TD}} + p_{1,29}^{3^{TD}} + p_{1,30}^{3^{TD}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \sum_{j=1}^{30} p_{1,j}^{3^{TD}} = 1
\end{bmatrix}$$

Finally, the MTM when the 'Turning' action is chosen ($\boldsymbol{P}_3$) is composed as follows:

$$\boldsymbol{P}_3 = \left[p_{i,j}^{3} \right] = \begin{bmatrix}
\underline{\substack{0 \\ km}} & \underline{\substack{10k \\ km}} & \cdots & \underline{\substack{350k \\ km}} & \underline{\substack{states \\ with\ \ damage}} & \\
\boldsymbol{P}_{TND}^{(70\times70)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \left| \substack{0 \\ km} \right. \\
\boldsymbol{P}_{TND}^{(70\times70)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \left| \substack{10k \\ km} \right. \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\boldsymbol{P}_{TND}^{(70\times70)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \left| \substack{350k \\ km} \right. \\
\boldsymbol{P}_{TD}^{(70\times70)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \left| \substack{states \\ with \\ damage} \right.
\end{bmatrix}^{(2590\times2590)}$$

- **Reward/cost function**

As this problem used a reward maximization function to derive the expected total discounted value rewards, the values used to represent the costs of the maintenance operations must be negative (Chadès et al., 2014). To derive the reward/cost function, a reward vector (q) for each action chosen (a = 1, 2, 3) is specified.

It is assumed that the 'Do nothing' action (a = 1) does not hold any operational cost. However, it is important to guarantee, due to the state space adopted constraints, that when the wheelset reaches states with diameter equal to the scrap diameter, kilometres since last turning/renewal (kst) of 350,000 km or damaged states, other option different from 'Do nothing' is chosen. This is done by

setting at these critical states cost values larger than the ones used in the remaining actions. For these states, it was assumed that values of 10,000 € should be assigned as follows:

$$q_i^1 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ q_{70}^1(s_{70}) \\ 0 \\ \vdots \\ 0 \\ q_{140}^1(s_{140}) \\ 0 \\ \vdots \\ 0 \\ q_{140+70}^1(s_{140+70}) \\ 0 \\ \vdots \\ 0 \\ q_{2450}^1(s_{2450}) \\ q_{2451}^1(s_{2451}) \\ \vdots \\ q_{2520}^1(s_{2520}) \\ q_{2521}^1(s_{2521}) \\ \vdots \\ q_{2590}^1(s_{2590}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -10000 \\ 0 \\ \vdots \\ 0 \\ -10000 \\ 0 \\ \vdots \\ 0 \\ -10000 \\ 0 \\ \vdots \\ 0 \\ -10000 \\ -10000 \\ \vdots \\ -10000 \\ -10000 \\ \vdots \\ -10000 \end{bmatrix} \begin{matrix} \\ \\ \\ \rightarrow Scrap \quad diameter \\ \\ \\ \\ \rightarrow Scrap \quad diameter \\ \\ \\ \\ \rightarrow Scrap \quad diameter \\ \\ \\ \\ \rightarrow Scrap \quad diameter \\ \left.\begin{matrix} \\ \\ \end{matrix}\right. \\ kst = 350k \quad km \\ \\ \left.\begin{matrix} \\ \\ \end{matrix}\right. \\ \left.\begin{matrix} \\ \\ \end{matrix}\right. \\ damage \\ \\ \end{matrix}$$

For the ‘Renewal’ action (a = 2), a value of -800 € is set, regardless of the state a wheelset is, hence, the reward vector is as follows:

$$q_i^2 = \begin{bmatrix} q_1^2(s_1) \\ \vdots \\ q_{2590}^2(s_{2590}) \end{bmatrix} = \begin{bmatrix} -800 \\ \vdots \\ -800 \end{bmatrix}$$

Turning a wheelset without damage (wheel states without damage) is set as having a cost of 50 € while doing turning for correcting a damaged wheelset (wheel states with damage) is set as having a cost of 150 €. However, there are some critical states where a ‘Renewal’ action is needed. Those cases are the ones when the scrap diameter is reached and, for the MDP does not ‘choose’ a

'Turning' action but instead a 'Renewal' action, they are 'penalized' with a cost of 10,000 €. Summing up, the reward vector for the 'Turning' action is as follows:

$$q_i^3 = \begin{bmatrix} -50 \\ \vdots \\ -50 \\ q_{70}^3(s_{70}) \\ -50 \\ \vdots \\ -50 \\ q_{140}^3(s_{140}) \\ -50 \\ \vdots \\ -50 \\ q_{140+70}^3(s_{140+70}) \\ -50 \\ \vdots \\ -50 \\ q_{2450}^1(s_{2450}) \\ q_{2451}^1(s_{2451}) \\ \vdots \\ q_{2520}^1(s_{2520}) \\ -150 \\ \vdots \\ -150 \end{bmatrix} = \begin{bmatrix} -50 \\ \vdots \\ -50 \\ -10000 \\ -50 \\ \vdots \\ -50 \\ -10000 \\ -50 \\ \vdots \\ -50 \\ -10000 \\ -50 \\ \vdots \\ -50 \\ -10000 \\ -10000 \\ \vdots \\ -10000 \\ -150 \\ \vdots \\ -150 \end{bmatrix} \begin{matrix} \\ \\ \\ \rightarrow Scrap \quad diameter \\ \\ \\ \\ \rightarrow Scrap \quad diameter \\ \\ \\ \\ \rightarrow Scrap \quad diameter \\ \\ \\ \\ \rightarrow Scrap \quad diameter \\ \lceil \\ \\ \\ \lfloor \\ \lceil \\ \\ \\ \lfloor \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ kst = 350k \quad km \\ \\ \\ \\ \\ damage \\ \\ \end{matrix}$$

- **Optimal policy**

An optimal policy is computed. The optimal policy of the decision process associated with maintaining a railway wheelset can then be organized in a graphic table for all states (damaged and undamaged) with the evolution of the kilometres since last turning (*kst*), as shown in Figure 50 (a) and (b).

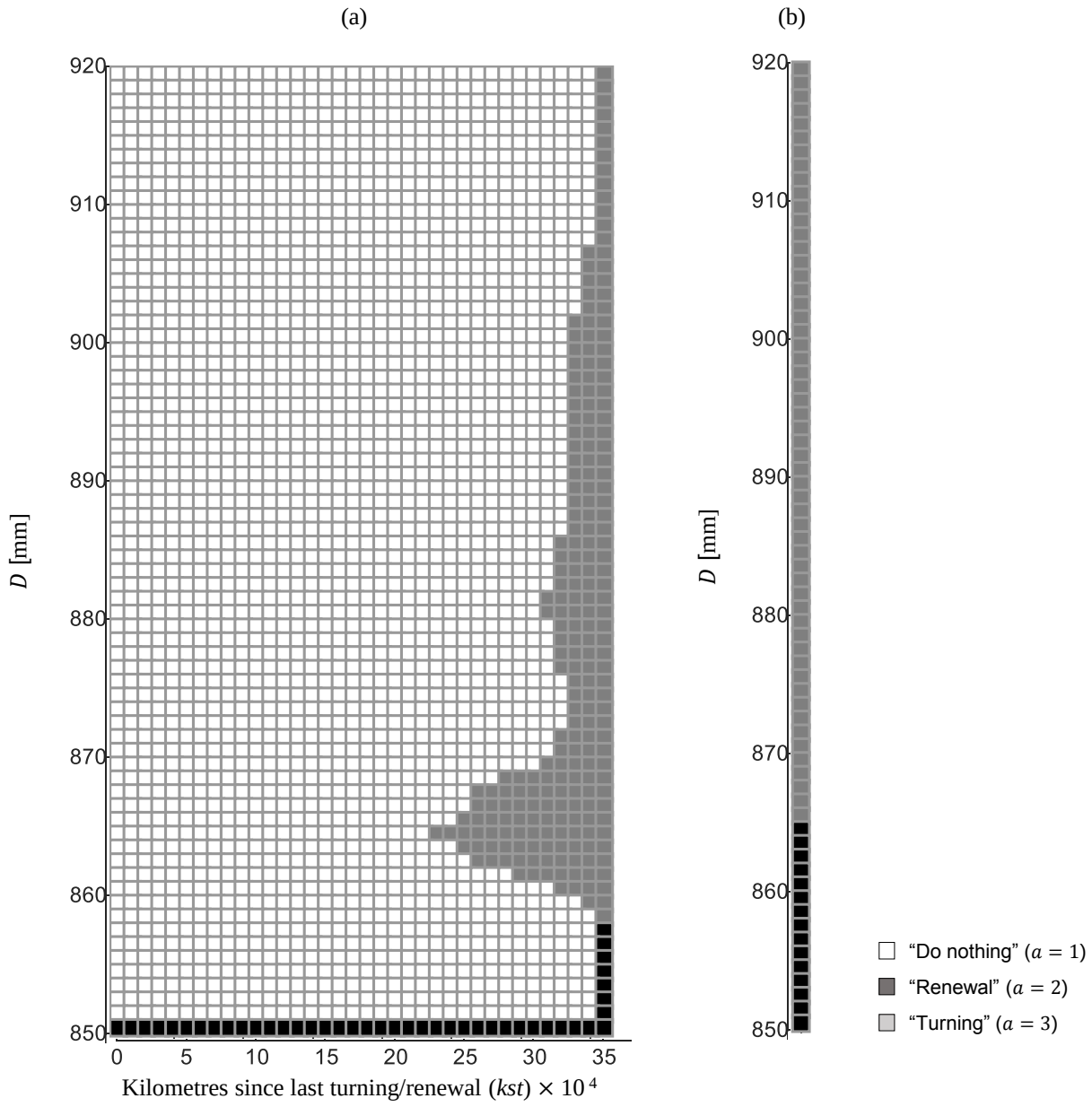


Figure 50: Map of decisions for wheelsets with (b) and without (a) damage with the evolution of the kilometres since last turning (kst).

By analysing Figure , one can see that the transition probability methods adopted and reward values were chosen in the section 'Reward/cost function' resulted in actions that were intended (a) for the undamaged wheelsets and (b) for damaged wheelsets. For the last, Figure (b) shows that only actions of 'Turning' or 'Renewal' are assigned, being the 'Renewal' actions recommended for the last states where the 'Turning' action would result in diameter values below the scrap diameter.

For the undamaged wheelsets, Figure 50 (a) indicates that the recommended actions change dynamically with kilometres since last turning (*kst*) and wheelset diameter (*D*). It helps to define a critical point, say k^*, as being the lowest *kst* such that turning action is recommended (here on coordinates *kst* = 240,000 km and *D* = 864 mm). The grey pattern, which covers part of the right side, suggests that turning actions should be performed earlier as the diameter decreases. This relationship holds until approximately the critical point k^*. For diameters below k^* (864mm or closer to the scrap diameter), the strategy shifts to allowing more *kst*. Indeed, for diameters below 858 mm, the best strategy is to not perform turning at all and let the wheelset wear out until scrap diameter. Hence, if one imagines a line connecting k^* to the first grey square on top (*kst* = 350,000 km and *D* = 920 mm) and a second line connecting k^* to the last grey square on bottom (*kst* = 350,000 km and *D* = 858 mm), it is evident that both signs and absolute values of the corresponding slopes are different. For newer wheelsets until about the point *D* = 870 mm, there is a slow decrease of *kst* while the decrease in *D* occurs at a higher rate, meaning that those wheels can support long periods without undergoing "Turning" actions. Then, from *D* = 870 mm down to k^* diameter, there is a fast decrease of both *kst* and *D*. Below k^* diameter, i.e. wheelsets whose diameters are close to scrap diameter, the "slope" changes sign and there is a fast increase of *kst* as *D* decreases.

Therefore, Figure 50 serves as a guideline for condition-based maintenance, that is, depending on the diameter (*D*), kilometres since last turning (*kst*) and whether or not damage has occurred, it provides the optimal strategy, i.e. the set of action that minimizes the total costs for each defined wheelset state. However, such optimal strategy/policy may not be effective in practice, as this requires train operating companies to have exceptional maintenance management and control over their assets, which might be unrealistic. Therefore, a modification of the policy, making it vary across only one parameter, in this case, kilometres since last turning (*kst*), can be compared with the optimal policy's expected cost so that an "easy-to-apply" maintenance strategy that has cost closest to the optimal can be implemented.

For this new strategy, some entries of the undamaged policy in Figure 50 (a) are modified according to the following rule: for a fixed "kilometres since last turning" (*kst*) value, all squares at that column or before that value's column will be filled with 1's (corresponding to action 'Do nothing'), all the squares after that value's column will be filled with 3's (corresponding to action 'Turning'), with the exception being the squares in black (action 'Renewal'), which will remain the same as in the optimal policy. Therefore, with the exception of the renewal squares, all the squares across the same value of "kilometres since last turning" (*kst*), i.e., all the entries on the same column, will have either 1 or 3, independently of the diameter (*D*). For example, if the new strategy sets the turning action to be performed after 150,000 *kst*, then all squares before and at column 150,000 *kst* will be filled with 1, or colour white, and all other squares will be filled with 3, or colour grey, (except the ones originally marked with 2, which remain black).

Under this framework, 21 different values of *kst* were used to build new policies and had their expected long-run cost extracted (no changes were made to the transition probability matrices or cost vectors, inputs of MDP approach). The cost results for these 21 different policies can be

compared with the long-run expected cost arising from the optimal policy displayed in Figure 51. To facilitate the comparisons, the optimal solution cost is set to 1, and all other policies' costs (which are higher) are displayed as percentages of increase compared to the optimal one, as shown in Figure 51.

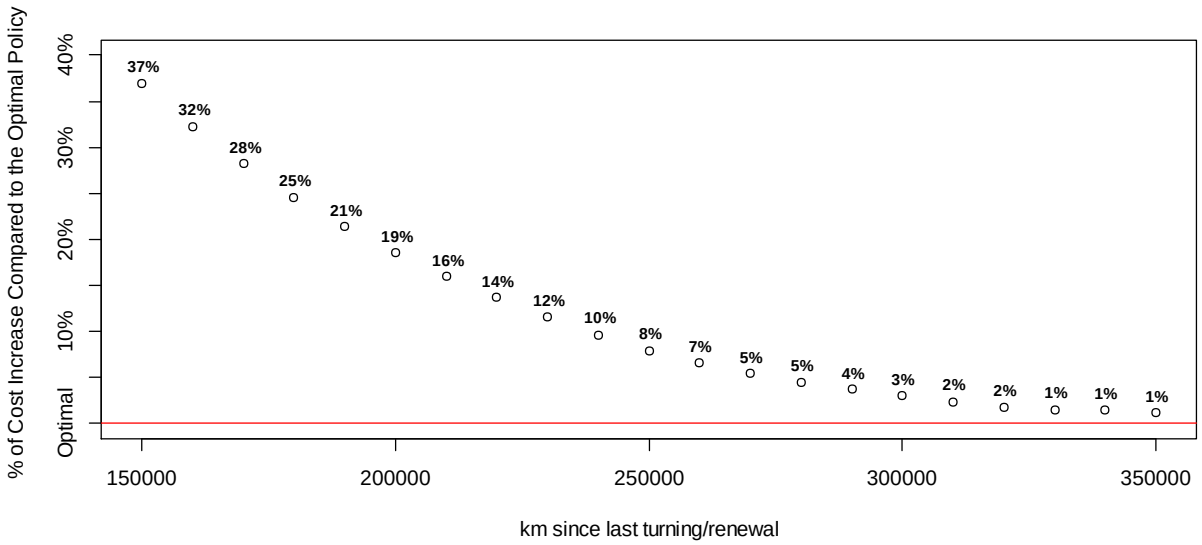


Figure 51: Comparison of optimal policy cost with costs arising from different policies based on 'kst' cut-offs.

Figure 51 compares the results of the different *kst* "cut-off" policies for 21 different values (empty circles) of kilometres since last turning, namely from 150,000km to 350,000km in steps of 10,000km, to the long-run average cost of the optimal policy (displayed as a red line). The optimal policy is a line, not a point, since it is dynamic in the sense that there's no kst "cut-off" as it considers a policy involving both *kst* and *D*. The optimal policy has the lowest expected long-run cost, followed by the policy which sets the "cut-off" for *kst* as 350,000km, with has an associated average cost about 1.1% higher than the optimal one. Policies with "cut-offs" in the neighbourhood of 350,000km, i.e., 330,000km and 340,000km, have also similar costs, about 1.5% and 1.4%, respectively, higher than the optimal. In general, policies with "cut-offs" inside the interval]300,000km, 350,000km] have associated long-run average cost within 2.5% of the optimal policy's cost and, hence, perform fairly good.

- **Main conclusions of the MDP approach**

A data-driven model based on the MDP approach was implemented in order to provide the train operating company with a better decision-making process in terms of wheelsets' turning and replacement policy. The change in tread diameter, kilometres since last turning and damage occurrence were used to define a discrete state space with a total of 2,590 states. A set of 3 possible actions were considered to account for all possible decisions that can be undertaken after a wheelset is measured, namely: (1) 'Do nothing', (2) 'Renewal' and (3) 'Turning'. Reasonable values for the cost vectors were set and optimization in terms of minimization of costs was performed with the support of the MDP Toolbox (Chadès et al., 2014).

The main result of this work is a dynamic map of decisions, in terms of actions to be taken for each discretization of kilometres since last turning (*kst*) and wheelset diameter (*D*). Identification of a critical point k^*, as tipping point for turning decisions, suggested three main trends: for wheelsets in earlier stages of life-cycle (until about 870mm of diameter), there is a slow decrease of *kst* as *D* decreases at a higher rate, hence, turning actions can be postponed to higher *kst* without the need of constant measurements of the diameter (the "slack" values for diameter lie within a big interval, in this case ranging from 920mm down to 870mm, or a 50mm slack). Then, from *D* = 870 mm down to k^* diameter of 864 mm, there is a fast decrease of both *kst* and *D*, so an efficient and more regular monitoring on the values of the diameter would be necessary to guarantee that turning actions would be performed at the optimal configuration presented in Figure 50 (in this case, the "slack" values are constrained to a small interval of only 6mm). Finally, for wheelsets which diameters are less than k^* diameter of 864 mm, or close to scrap diameter, there is a fast increase of *kst* as *D* decreases, and the recommendation is to allow the wheelset to wear out until scrap diameter, as many frequent inspections would be necessary to guarantee that turning action would still be captured in an "optimal" setting, and hence, it is unlikely that turning at this point would be cost-effective.

Another contribution of this work concerns the ease of policy implementation. As a dynamic policy contemplating two parameters might be hard to implement in the context of a train operating company, it is interesting to compare the optimal results with those obtained from simpler policies. Hence, policies based on 21 different "cut-offs" values for kilometres since last turning were tested so that, for a given value of the "cut-off" (e.g. 250,000km), all actions before or on that mark would correspond to 'do nothing' and all actions after that mark would result in 'Turning'. The exception would be the actions assigned 'Renewal' in the optimal policy, which were kept the same. In this framework, no consideration of the diameter would have to be made by the maintenance team in order to decide between 'Turning' or 'Do nothing', as *kst* would be the only input needed. Results displayed in Figure 51 revealed that, although the optimal solution had the lowest expected long-run cost, a good strategy would be to set the "cut-off" for *kst* as close as possible to 350,000km, independently of the diameter. This strategy would increase the average long-run cost by only about 1.1%. In general, policies with "cut-offs" within the interval]300,000km, 350,000km] perform fairly good and are recommended.

By inspecting the different cost values in Figure 51, it is not clear that, in the presence of states for which *kst* would exceed 350,000km, the optimal configuration would lead to higher values of *kst* associated with turning recommendation. In fact, considering the train operating company's policy, even the recommendation of turning at 350,000km translates into doubling the amount of *kst* that is currently established in the maintenance strategy. In this context, investigating further the *kst* of 350,000km may be unrealistic, hence the limited state space based on maximum *kst* of 350,000km is taken as one of the limitations of this study. As a second limitation, although the state space described a range of different configurations involving kilometres since last turning, wheel diameter and occurrence of damage, it did not control for the evolution of other important variables such as the flange thickness and height as well as the angle inclination. Inclusion of these variables is

suggested as a future step, as the current technical standard imposes limits to these additional parameters.

3.4 TACTICAL MAINTENANCE PLANNING

This section presents the application of the tactical maintenance planning model to the case study of Fertagus. First, some details on the Fertagus train operating company are provided and the second subsection presents problem specifications and values for all parameters. Cost values are given in monetary units for the sake of confidentiality. Therefore, in order to convert to €, a multiplication coefficient between 1 and 10 must be used.

Fertagus train operating company

Fertagus trains run on a line of 54 kilometres that cross the “25 de Abril” bridge; and stop at 14 stations. Total travel duration between Roma-Areeiro (in Lisbon) and Setúbal is 57 minutes. No train can be pulled out of service to go to maintenance if there is no backup train available, and thus Fertagus owns 18 train units, when only 17 are necessary to perform the current operation. The question of whether or not this could be done differently is out of the scope of the present work and is left for further research.

Fertagus maintenance yard comprises a total of 10 lines, though they are numbered from 1 to 12, as lines number 3 and 4 were never built but were in the original design of the maintenance yard. Out of these 10 lines, only 3 are used to perform maintenance activities, respectively lines 10, 11 and 12. The other lines are used for testing or parking. Table 10 provides information on the use that is given to each line in Fertagus maintenance yard.

Table 10: Use of the lines in Fertagus maintenance yard

Line numbers	Use in the maintenance yard
1 - 2	Several tests
5 - 8	Parking
9	Cleaning operations and conservation cleaning
10 - 11	Maintenance with catenary
12	Maintenance without catenary (for pantograph replacement)

All lines used to perform maintenance activities (lines 10, 11 and 12) are indoors. On the contrary, line number 9 which is dedicated to cleaning is not covered because this maintenance activity is to be performed outside. Fertagus maintenance yard also performs wheelset turning within the maintenance yard in an underfloor wheel lathe.

- **Specific input parameters**

In order to model the case study, information on Fertagus maintenance activities was gathered in order to have the correct inputs for the parameters of the mathematical model, through meeting, e-mails and phone calls. The maintenance activities planned by Fertagus are summarized in Table 11.

Table 11: Maintenance activities planned in Fertagus case study.

Maintenance activity	Activities performed	Period	Time needed	Work force required
ETS	Mostly inspection activities	Every 5 weeks	1.5 – 2 h	4
VEq	Inspection of motor block, pressure check, etc.	Every 37,500 km	6 h	4
VP	Doors check-up	Every 150,000 km	6 h	4
VL	Lubrication check-up	Every 120,000 km	4 – 6 h	4
VEI	Electric system check-up	Once a year (before Winter)	40 h*	4
VS	Biannual check-up	Twice a year (beginning of Spring and end of Summer)	12 h	4
TRF	Wheelsets turning	Every 120,000 km	16 h	2
V1	Some parts of the pantograph are maintained	Every 300,000 km	-	4

* Not continuous, i.e. maintenance task is split in different days.

The parameters of the mathematical model come from the Portuguese train operating company (Fertagus) and they can be found in Tables 11 – 14. There are eighteen trains so U is a set of integers from 1 to 18; there are sixteen maintenance activities that can be performed in Fertagus maintenance yard, which implies that I is a set of integers from 1 to 16; the planning horizon of the maintenance planning is set to a year, and since the time unit is a week, T is a set of integers from 1 to 53; and finally, four different spare parts are stored in Fertagus maintenance yard so P is a set of integers from 1 to 4.

In Table 12, all parameters depending on the maintenance activities i are summarized. The first line includes the name of maintenance activity i_1 (ETS), its cost is 614,42 monetary units. Then, the period of the ETS maintenance is displayed in weeks and is equal to five weeks. This means that maintenance activity i_1 is due every five weeks. Then, both the working load and the duration of the maintenance activity i1 are given. ETS maintenance is a 10 person-hours maintenance activity and lasts 2,5 hours long. Finally, the set of maintenance lines where maintenance activity i_1 can be performed is displayed, i.e. it can be performed either on line 11 or on line 12 of Fertagus maintenance yard. For instance, wheelset turning (TRF) is maintenance task number 5 (i_5).

Parameters that depend on the spare parts p are displayed in Table 14. In the first line, information on the spare part $p = 1$ is given: its name, its cost per week, the number of weeks needed to maintain the spare part, and the maximum number of spare parts according to the maintenance yard storage area.

Table 13 provides information on the initial conditions of Fertagus train units, i.e. the time intervals (in weeks) between the last maintenance activity i and the beginning of the planning horizon for train unit u. It must be highlighted that all values of $O(u, i)$ can be deduced from Table 13. For example, $O(u_2, i_3)$ which corresponds to the last time maintenance activity i_3 was performed on train u_2, is set to 23 weeks, meaning that maintenance activity i_3 was last time conducted 23 weeks before the beginning of the planning horizon.

Table 12: Parameters for each maintenance activity i in Fertagus case study.

Maintenance activity (i)	Cost of maintenance activity $C(i)$	Period (in weeks) $T(i)$	Amount of work (in hours) $f(i)$	Duration (in hours) $d(i)$	Lines' set $L(i)$
1	614.42	5	10	2.5	{10, 11}
2	1,720.37	16	28	7	{10, 11}
3	1,018.98	53	14	3.5	{10, 11}
4	829.17	63	14	3.5	{10, 11}
5	2,522.22	63	42	21	{10, 11}
6	815.28	63	14	4.6	{10, 11}
7 - 11	3,516.44	53	12.4	3.1	{10, 11}
12	793.29	53	14	7	{10, 11}
13 - 14	396.64	53	3.5	0.88	{12}
15	56.25	26	1	1	{10, 11}
16	2,457.68	136	40	10	{12}

Table 13: Time interval (in weeks) between last maintenance activity I and beginning of the planning horizon for train unit u, for Fertagus case study.

Train unit (u)	Maintenance activity (i)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	4	3	19	50	46	25	40	40	40	40	40	13	16	16	19	45
2	2	5	26	7	42	25	40	40	40	40	40	13	16	16	9	40
3	4	0	37	19	20	24	36	36	36	36	36	12	16	16	16	54
4	3	10	40	30	15	20	35	35	35	35	35	12	16	16	18	56
5	3	15	16	16	15	24	15	15	15	15	15	12	15	15	24	56
6	0	14	2	42	36	23	14	14	14	14	14	11	15	15	3	54
7	3	15	18	23	49	23	13	13	13	13	13	11	15	15	17	47
8	0	1	30	30	51	22	12	12	12	12	12	10	14	14	4	55
9	1	14	38	14	47	22	12	12	12	12	12	10	14	14	11	46
10	1	0	5	37	39	21	11	11	11	11	11	9	14	14	0	35
11	3	9	43	16	44	21	10	10	10	10	10	9	13	13	5	41
12	4	11	52	45	43	20	7	7	7	7	7	8	13	13	6	35
13	2	10	17	3	45	20	6	6	6	6	6	8	13	13	18	41
14	2	14	13	48	46	19	5	5	5	5	5	7	12	12	16	35
15	4	6	33	22	40	19	4	4	4	4	4	7	12	12	8	49
16	0	1	46	45	36	18	3	3	3	3	3	6	12	12	23	37
17	3	15	39	55	48	18	2	2	2	2	2	6	11	11	20	43
18	4	12	28	35	35	17	1	1	1	1	1	5	11	11	14	44

Table 14. Parameters for each spare part p in Fertagus case study.

Spare part p	Spare part type	Cost of having spare time per week $P(p)$	Maintenance Duration (in weeks) $R(p)$	Maximum amount of spare parts $A(p)$
1	Wheelset	104.17	1	20
2	Trailer bogie	1,041.67	0	20
3	Motor bogie	1,041.67	1	20
4	Pantograph	416.67	2	20

Regarding the values of parameter $q(i,p)$, i.e. the number of spare parts p required to perform maintenance activity i, as the majority of sixteen maintenance activities do not require spare parts, the majority of values are zero. However, some maintenance activities of Fertagus railway operating company still need spare parts, namely: $q(5,1) = q(16,2) = q(16,3) = q(16,4) = 1$.

Finally, all constants of the mathematical model for the Fertagus case study are provided. The planning horizon H is equal to 53 weeks. The shunting cost is set to 5 000 monetary units (this value was initially given as an approximation and will be subject to a sensitivity analysis in next section). The maximal working load k is 160 person-hours, and is calculated as the product of the number of persons working in Fertagus maintenance yard by the number of working hours per day times the number of useful days of the week, i.e. 4 persons × 8 hours × 5 (days) = 160 person-hours. The maximal working time per week is 40 hours and is calculated as the product of the number of working hours per day times the number of useful days in a week, i.e. 8 hours × 5 (days) = 40 hours.

Results and discussion

This subsection analyses the results of running the model for a small size case study of a Portuguese train operating company (Fertagus) described in Section 4. Firstly, it starts by analysing the optimality gap over computational time for the best solutions found, showing that after a computational time of 1 hour an acceptable optimality gap of 0.63% is achieved. Then, it presents the objective function and the distribution of the different cost components (A, B, C and D) of the objective functions for the optimal solution after 1 hour of computation. Finally, last subsections provide a sensitivity analysis of the shunting cost component and the maximal working time per week.

- **Analysis of the optimality gap versus the computation time**

As the size of the problem increases, the computational time to obtain the optimal solution also increases. However, a feasible solution can always be found within few minutes, with an optimality gap. The closer the optimality gap is to zero, the better the solution is. In the explored case study, it is interesting to study the evolution of the optimality gap with respect to computational time in order to know when the solution can be considered “acceptable” (i.e. lower than a certain optimality gap).

In this analysis, the computational time varied from 1 minute to 24 hours: 1 min, 5 min, 10 min, 15 min, 30 min, 45 min, 1 h, 1.5 h, 5 h, 10 h and 24 h. Figure 52 shows the graph of the optimality gap over the computational time. This graph shows that the longer the computational time, the smaller the optimality gap, though this decrease is not linear. The goal of this analysis is to be able to select the smallest computational time whose corresponding optimality gap is acceptable. It can be seen on the graph that after a calculation time of one hour (1 h), the optimality gap around 0,63% is achieved, and when the computational time is increased to one day (24 h), the optimality gap becomes slightly less than 0,60%, which is better (or lower), but may not be worth of the additional time spent to minimize the cost. Therefore, the computational time was chosen to be set to one hour for all further analysis in this section. Table 15 exhibits the detailed values of the computational time and the associated optimality gap.

Table 15. Values of computational time and corresponding optimality gap.

Computation time (s)	Optimality gap (%)
60 (1 min)	2.67
300 (5 min)	2.67
600 (10 min)	2.30
900 (15 min)	1.55
1,800 (30 min)	0.97
2,700 (45 min)	0.97
3,600 (1 h)	0.63
5,400 (1.5 h)	0.63
18,000 (5 h)	0.63
36,000 (10 h)	0.63
86,400 (24 h)	0.60

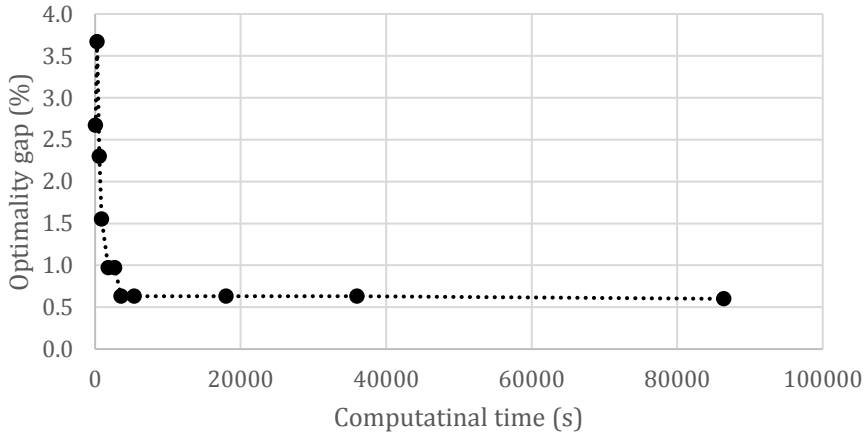


Figure 52: Optimality gap (%) with respect to computational time (s).

- **Cost components for the optimal value of the objective function**

After one hour of computation, the best solution found has a value for the objective function equal to 1,664,750 monetary units. This total cost is divided into the four cost components A, B, C and D whose values are displayed in Table 16.

Table 16. Distribution of the different cost components in the objective function (after 1 hour of computation)

Cost component	Value (monetary units)	Percentage of the total cost (%)
A (maintenance activities cost)	670,495.00	39.89
B (shunting cost)	1,005,000.00	59.78
C (spare part cost)	5,521.01	0.33
D (penalty cost)	15.50	0.0009

Note that the cost component B is the one with the largest impact on the total preventive cost which is why maintenance activities are grouped whenever it is possible. The cost component A, which is the maintenance activity cost, has the second largest impact. However, it is not a cost that is easily changed as it is dependent on the maintenance activities themselves. The cost components C and D have smaller influence on the objective function.

This section conducts sensitivity analysis to two main parameters of the model: i) the shunting cost and ii) the working time per week. Note that changing the shunting cost will only affect the parameter S, whereas changing the working time per week will affect both the maximum working load (k) and the maximum working time (w).

The variation of the objective function compared with the reference case ($S = 5{,}000$) was calculated, considering a computational time of 1 hour. Moreover, the ratio between the objective function variation and the shunting cost variation was also computed and it is presented in the last column in Table 17. The value for a variation of zero percent is not presented as a division by zero would be involved.

For example, a shunting cost variation of +2.0% induces an objective function variation of +1.3%, which corresponds to a ratio of 0.65 between the variations. It is interesting to notice that the objective function variations exhibit values nearly anti-symmetric, meaning that an increase of 6% of the shunting cost component will induce a raise on the objective function of around 3.58%; and a similar decrease of 6% in the shunting cost component will also induce a decrease of the objective function of around 3.54% (i.e. -3.54% of variation). Nevertheless, it must be said that the values induced by a negative shunting cost variation are always slightly higher than the ones induced by a positive shunting cost variation. For instance, a variation of the shunting cost of -2% induces a total cost variation of 0.90% while a variation of shunting cost of +2% induces a total cost variation of 1.30%.

Table 17: Sensitivity analysis of the shunting cost and relation between two variations.

Shunting cost (S)	Shunting cost variation (%)	Objective function variation (%)	Ratio between objective function variation and the shunting cost variation
4,500	-10.0	-5.43	0.54
4,600	-8.0	-4.12	0.51
4,700	-6.0	-3.54	0.59
4,800	-4.0	-2.63	0.66
4,900	-2.0	-0.90	0.45
5,000	0	0	-
5,100	+2.0	+1.30	0.65
5,200	+4.0	+2.97	0.74
5,300	+6.0	+3.58	0.60
5,400	+8.0	+4.46	0.56
5,500	+10.0	+5.68	0.57

On the working time per week. Current preventive maintenance shifts in Fertagus are of 8 hours per day, and five days a week; which corresponds to a current maximum working time in the maintenance yard of 160 hours per week, assuming that the preventive maintenance team includes 4 persons. In order to quantify the impact of the time allocated to preventive maintenance, a sensitivity analysis is conducted on the working time per week. It must be highlighted that changing the working time per week affects two parameters of the mathematical model which are k and w. The time allocated varies from 36 hours to 44 hours; in which 40 is the reference (all values can be found in Table 3.18). The allocated person-hours per week and allocated hours per week are dependent through the following relation: person-hours per week = (hours per week) × (number of persons). In this sensitivity analysis, the number of persons did not change. When the time allocated is 36 hours, it finds no feasible solution. On the other extreme, i.e. when the time allocated is 44 hours, an optimal solution is found in 2,445.8 seconds (40 min and 45 seconds) with an optimality gap of 0%, which means that the calculus stops before the end of the reference computational time (1 hour). For an allocated time of 42 hours, the solution is also optimal after 2 684 seconds (44 min and 46 seconds) with an optimality gap of 0%. There are two possible explanations for this; either 160 person-hours working load maybe considered optimal in Fertagus case study; or the initial conditions have a lot of influence on the maintenance planning. Indeed, a maintenance which is done by four working persons for several planning horizons could lead to initial conditions that require a preventive maintenance done by four persons.

Table 18: Sensitivity analysis of the shunting cost and relation between two variations.

Allocated hours per week	Allocated person-hours per week	Optimality gap after 1h computation (%)	Total maintenance cost (monetary units)
36	144	No feasible solution	
38	152	1.84	1,701,030
40	160	0.63	1,675,360
42	168	0	1,659,800
44	176	0	1,659,750

Optimizing total costs of preventive maintenance in a train fleet is the aim of the present research work, namely by creating a mixed-integer linear programming model for a train operating company. The mathematical model was explained in detail and an illustrative example was provided to improve reader's understanding of the model. Then, it was applied to the case study of Fertagus train operating company. Details on the maintenance yard configuration (e.g. lines) were discussed and two new constraints were added, which represent a contribution to current models (Doganay and Bohlin 2010, Bohlin and Wärja 2015). Data for the Fertagus case study was collected, in which costs are presented in monetary units for the sake of confidentially, though they maintain their relative weight between each other. Although, the final mathematical model was adapted to the specific case study of Fertagus train operating company, it can very easily be modified to fit to any train operating company's specifications. Results suggest that after a computational time of one hour, the best solution found has an optimality gap of 0,63%, which is considered acceptable. Sensitivity analysis on the shunting cost and on the working hours per week were conducted, showing that the shunting cost has an important effect in the objective function, and that increasing the working hours per week provide even optimal solutions, obtained in less than 1 hour of computation.

This mathematical model enables to find an optimal maintenance planning but it is of course user input dependent and this is a limitation in any kind of these models. Indeed, if the user inputs do not represent correctly the real-life situation, the maintenance planning could hardly be optimal. For the Fertagus case study, corrective maintenance activities were still not included in this model, though it should be emphasized that Fertagus train operating company has a dedicated team for corrective maintenance that works in another shift than the preventive maintenance team. Optimizing preventive and corrective maintenance activities together might bring better results than optimizing each one separately, though it would increase the size and complexity of the model.
Moreover, the cost component D has to be improved to be transformed to monetary units. Note that the objective function was adapted from Bohlin and Wärja (2015). Although in practical terms, it prevents the model to maintenance too early, and thus it prevents from losing life of the component by reducing the maintenance cycle. This loss of residual life should be expressed in monetary units. It is important to mention that the present model does not take into account the availability constraints explicitly, and this is a current limitation of the present model. Although in the Fertagus case study, there are more trains than the ones needed to perform the train operations in the daily schedule, an operational planning model is still needed to be integrated (with a horizon of a week and the modelling of train movements inside the maintenance yard and in the network).

3.5 OPERATIONAL MAINTENANCE SCHEDULING

This section 3.5 applies the operational maintenance scheduling model to the Portuguese train operating company Fertagus. Again, the train operating company is responsible to operate the railway service and for the safety and maintenance of the trains. Their fleet is composed of 18 trains, from which 17 train units are available for operational services and are intended to cover the set of tasks and preventive maintenance actions. The application of the model to this case study has the main goal of finding the best feasible solution that outputs a rolling-stock plan for a planning period of two days (48-hour period). A representative period was chosen to perform this study.

Appendix B contains all the information needed to run the model for the case study of a Portuguese train operating company, namely tables B.1 to B.8 provide values for the parameters used in this case study. Table B.1 provides stations name, their corresponding number and their associated minimal turning time (in minutes). Table B.2 provides information on the parameter $W(s,s')$, i.e. the possible dead-headings between two stations: if the value equals zero, a dead-heading between stations s and s' is not possible; otherwise, its value would be equal to one and a dead-heading is possible. Then, Table B.3 provides the distances/lengths (in kilometres) between stations and Table B.4 provides the associated durations (in minutes) between stations. Table B.5 provides the model constants and Table B.6 provides a small excerpt of the table with all train service tasks. The complete table is not presented but is provided in the electronic supplementary material. Table B.6 identifies the various tasks on the first column. The next columns give the required number of units, the maximal number of units, the departure station, the arrival station, the departure time and arrival time of a task. Tasks 1 to 790 are real tasks. The last 30 tasks are beginning and ending virtual tasks. Table B.7 provides information on the maintenance tasks that need to be performed in the planning period through the parameter $KM_{k,m}$, i.e. if it is equal to 1, then maintenance task m must be scheduled for train unit k. Finally, Table B.8 provides information on the maintenance durations and amount of work (work load) for each maintenance task.

The optimal minimum cost is 150128 and it takes 1711.70 seconds (less than half an hour) to achieve the optimal solution, which provides a schedule of all railway operations and maintenance tasks. The optimal solution provided a different train unit assignment that is currently implemented in the company. One aspect that the optimized solution showed is that there is another way to assign train units to normal train services, while saving a significant amount in the total distance run in dead-headings. Table 19 provides a comparison of the total dead-headings for the case study between the current situation and the optimized by the application of the model. It shows that the model allows a reduction of the total distance run in dead-headings (206 km per day to 131 km per day).

Table 19 – Comparison of the total dead-headings for the case study between the current situation and the optimized by the application of the model.

Pairs of Stations	Distance (km)	Current situation	Model application
Roma-Areeiro – Pragal	11.68	4	4
Roma-Areeiro – PMC	25.60	2	0
Coina – PMC	1.70	14	16
Setúbal – PMC	28.60	2	2
Coina - Setúbal	26.83	1	0
Total distance (km)		206	131

- **Reliability analysis using uncertain durations for maintenance tasks**

This section explores the role of uncertainty in the duration of maintenance tasks, providing a reliability analysis of maintenance schedules within the train service operations. This reliability analysis has the main objective of assessing the probability of achieving a feasible schedule, given the uncertainty associated with maintenance durations. Therefore, the decision maker (the maintenance planner) can assess how reliable a given maintenance schedule/plan is and what is the impact of the uncertainty in the durations of maintenance tasks in the value of the objective function. In the previous example, maintenance durations were assumed to be deterministic (constants that are known in advance). In this subsection, they will be modelled as random variables to account for their uncertainty, by assigning probability density functions to each one of them: MT1, MT2, MT3, …, MT14 from Table A.8. Inspired by common practice in Project Evaluation Research Technique (PERT) applications, a generalization of a Beta probability distribution is assumed as the distribution of each maintenance duration. Table 20 provides the parameters of the probability distributions for the durations of each maintenance task. Estimates for the PERT parameters were achieved through interviews and using expert judgement techniques to maintenance planners/workers.

Table 20 – Parameters of the PERT probability density functions assigned to each random variable (MT_m): minimum (a), most likely (b) and maximum (c) parameters.

Random variables	Parameters of PERT distributions		
	Minimum (a)	Most likely (b)	Maximum (c)
MT_1	100	150	200
MT_2	300	420	600
MT_3	150	210	300
MT_4	150	210	300
MT_5	200	276	400
MT_6	120	186	280
MT_7	120	186	280
MT_8	120	186	280
MT_9	120	186	280
MT_{10}	120	186	280
MT_{11}	300	420	600
MT_{12}	40	53	90
MT_{13}	40	53	90
MT_{14}	45	60	90

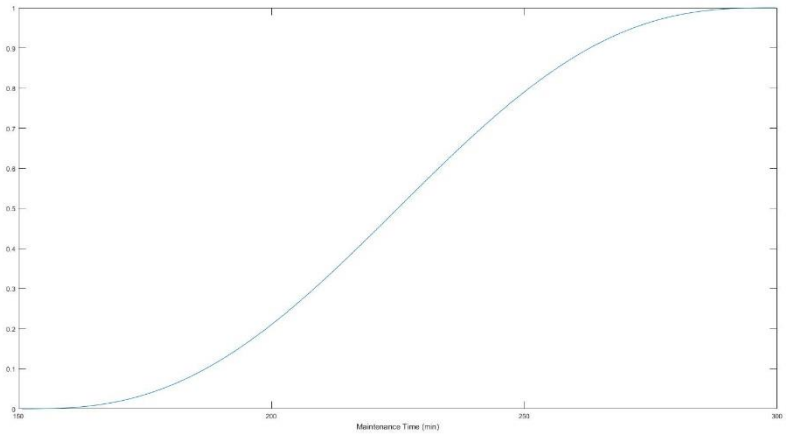


Figure 53: An example of the cumulative probability function for MT3 (duration of maintenance task m=3), which is assigned a PERT probability function with parameters a=150, b=210 and c=300.

Note that the choice of bounded probability density functions allows the computation of the worst and best scenarios. The worst scenario corresponds to the scenario with higher durations for each maintenance task, i.e. $MT_m = c_m$ for all maintenance tasks. The best scenario corresponds to the scenario with lower duration for each maintenance task, i.e. $MT_m = a_m$ for all maintenance tasks. The case study explored was run assuming that each maintenance task has its most likely duration, i.e. $MT_m = b_m$ for all maintenance tasks.

A design of computer experiments was conducted with the following steps: i) running the worst and best scenarios and assess the optimal objective value; ii) running a set of simulated experiments (N=60 simulations) using a Monte Carlo simulation procedure to generate random values for the maintenance durations for each maintenance task; iii) compute the empirical cumulative distribution of the optimal value for the objective function and iv) assess the probability of achieving a feasible schedule and the probability that a higher value than the optimal value for the best scenario is achieved.

Figure 54 provides the empirical cumulative distribution of the optimal value for the objective function for feasible solutions, exhibiting no increase in the objective function for 36.7% (22 out of 60) of the simulated cases and the maximum increase of +84 in the objective value from 150128 to 150212. Table 21 provides the main results from the computer experiments run, namely for the extreme cases (best and worst scenarios), a 'most likely' scenario corresponding to the case where all maintenance durations are equal to the most likely duration for each maintenance task (i.e. $MT_m = b_m$ for all maintenance tasks) and an average case which takes the average of all computed cases.

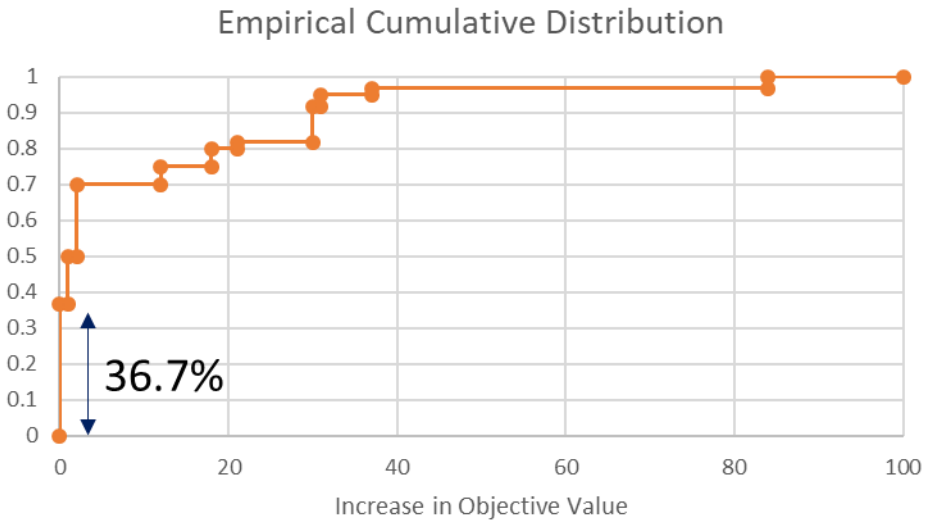


Figure 54: Empirical cumulative distribution of the optimal value for the objective function.

Regarding the extreme scenarios, the best scenario revealed no increase in the objective function, whereas in the worst scenario, the high maintenance durations made the problem unfeasible. From the design of experiments, it was shown that only the worst scenario revealed an unfeasible solution, which indicates that the expected probability of achieving an unfeasible solution should be lower than 1/60. In fact, from our initial tests, it is very likely that it is much lower than 1/60 as the PERT distributions assign very low probabilities to extreme values (a, c). In further research, a rare-event simulation procedure within an adaptive design of experiments will be defined, in order to improve the way this probability is assessed using less computational time needed.

Table 21: Objective Function values and computational times for the 'most likely' scenario, the 'best' scenario, the 'worst' scenario and the average case.

Simulations	Objective Function value	Computational time (s)
'Most likely' scenario	150128.00	1711.70
'Best' scenario	150128.00	2492.20
'Worst' scenario	No feasible solution	27.0
Average	150137.83	1102.16

The present section explored the integration of maintenance tasks and their associated constraints into a fleet assignment model to schedule train services. An Integer Linear Programming (ILP) was introduced, based on the previous work of Tréfond et al. (2017). The final model is an ILP that assigns train units to operational services as well as maintenance tasks in the time interval between operational services. The proposed ILP model is tested for an illustrative example and for a case study, and it was shown that the proposed model is able to find optimal solutions within reasonable time (less than an hour of computational time), providing a schedule for maintenance and availability of operations in a train operating company. Moreover, a reliability analysis based on a design of computer experiments was conducted to show the influence of the uncertainty associated with the time durations of maintenance tasks. It was shown that for the tested cases, the uncertainty

associated with maintenance durations did not influence dramatically the optimal value for the objective function, and thus, it did not influence the optimal schedule. Nevertheless, for the worst scenario no feasible solution was found which means that no feasible schedule exists, and thus, under the worst scenario, the high maintenance durations will affect the normal operations and cause delays. This method to assess the impact of the uncertainty in maintenance durations is used to understand the potential impact in the optimal value of the objective function.

3.6 UNCERTAINTY ASSOCIATED WITH INSPECTION

Data acquisition plays a substantial role in determining cost-effective maintenance policies for railway wheelsets. Wheelsets deteriorate through time, and a good maintenance strategy should keep their condition at an acceptable level by monitoring various important parameters and ensuring they are under control. At the same time, it is crucial to accurately measure those parameters, so that turning and replacement actions can be undertaken very closely to cost-effective targets, where costs of losing wheels' useful life are minimal. In this section, a comparison between data acquired from manual (gauge device), laser and turning measurements is presented for three wheelset parameters: Flange Thickness (F_t), Flange Height (F_h) and Flange Slope (q_R). The main interest lies on the comparison of precision of these measurements: intuitively the distribution of each type of measurement (for each parameter) is centred at some target value (which should be approximately the same across the different types), and it is desired to know whether the variance around the target value is different for manual, laser and/or turning measurements. Although the three technologies have not been used complementary, i.e., the wheelsets were not measured by the three of them at the same time, if an accurate model can be obtained to successfully explain how the statistics change over some distance measure (e.g., kilometres since last turning/renewal), then the task can be reduced to an analysis of the residuals, which is performed in this case study.

To motivate the research question with a visual example, assume that the distributions of the parameters for each of the three measurement types is approximately normal, centred at some target value, but with different variances. Then the expected behaviour of the distributions should look like the one displayed on Figure 55:

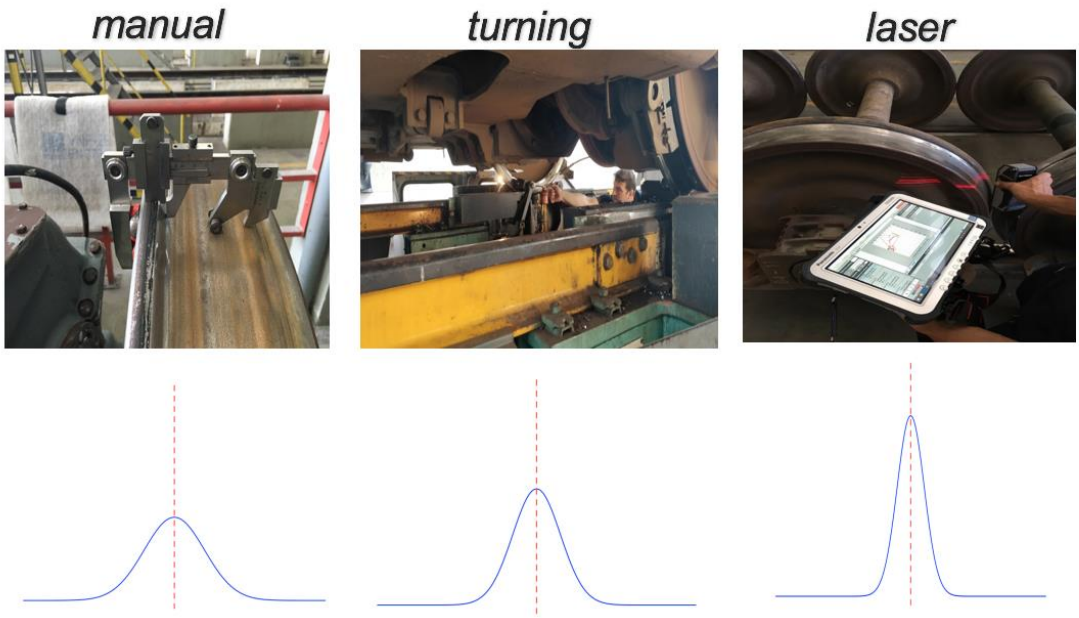


Figure 55: Example of hypothetical distributions for different measurement types.

If data exhibited the behaviour on Figure 55, it would be possible to quantify, in terms of financial savings, the differences of adopting one technology in lieu of the other(s), i.e., it would be possible to answer the following question: can a better precision (less variance) bring savings on maintenance costs? The goal of this section is to investigate the actual behavior of data to see if there is evidence that allows for this kind of inference regarding the variances of different types of measurements.

Data was again acquired from Fertagus that transports passengers in a single line, which extends 54 kilometers and servers 14 stations. A 17-years interval database, ranging from January 2001 up to November 2018, is considered. The company operates 18 trains (units). Each unit has 4 cars, and each car has eight wheels (i.e. four wheelsets). Figure 56 provides a schematic representation of a four-cars unit.

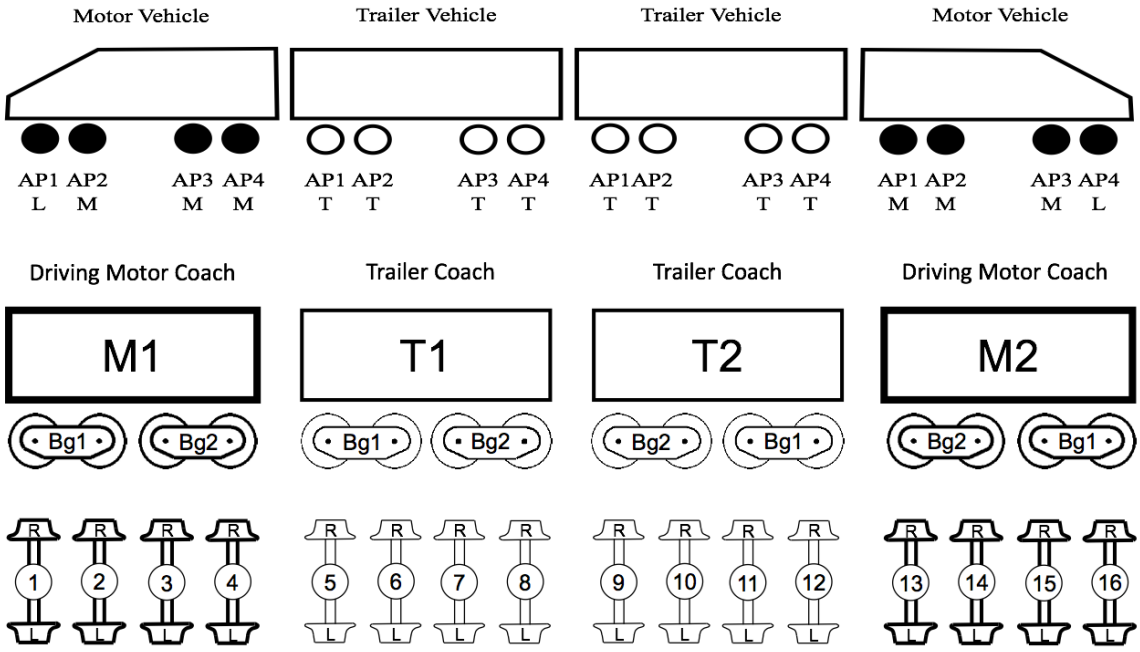


Figure 56: Schematic representation of a 4-cars unit. Source: the authors.

One parentheses here is that, for this study, instead of considering the 18 unit numbers with the above characteristics, 36 unit numbers were considered, depending on the wheelset's position, i.e., the serial numbers 3501 up to 3518 accounted for the first two vehicles (M1 and T1, or wheelsets 1-8), and the serial numbers 3551 to 3568 accounted for the last two vehicles (T2 and M2, or wheelsets 9-16). The reader should notice the equivalence between serial numbers 3501 and 3551, or 3502 and 3552, and so on: they correspond to the same unit, but different vehicles within the unit.

The database contained the following information: unit number, unit running kilometers (cumulative kilometers), measurement type (manual, laser or turning), vehicle type (leader, motor or trailer), date, wheelset position (1 up to 16), side of wheel (left or right), technician responsible for the measurement, renewal cycle (H1 or H2, with notable differences in the hardness of wheels from

one cycle in comparison to the other), Flange Thickness (F_t), Flange Height (F_h) and Flange Slope (q_R). Figure 57 provides a schematic representation of the flange thickness and the flange height measures and the tread datum position (70 mm measured from the flange back and A = 13 mm for UK profiles or A = 10 mm for EN profiles).

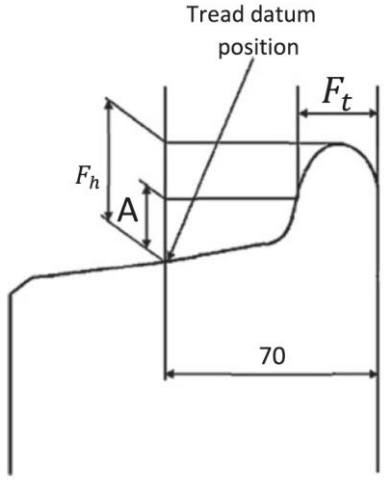


Figure 57: Flange Height F_h and Flange Thickness F_t. Source: adapted from Andrade & Stow (2016).

From this original database, it was possible to compute other variables of interest: kst (kilometers since last turning), ΔF_t (change in flange height since last turning due to wear), ΔF_h (change in flange height since last turning due to wear), Δq_R (change in flange slope since last turning due to wear). Table 22 provides an overview of the aforementioned variables as well as some relevant descriptive statistics.

Table 22: Main variables and descriptive statistics.

Variable	Description	Type	Mean	Min	Max
ΔF_t	Change in the flange thickness (in mm) since last turning due to wear	Continuous	-0.382	-9.2	8.0
ΔF_h	Change in the flange height (in mm) since last turning due to wear	Continuous	1.50	-7.9	8.0
Δq_R	Change in the flange slope (in mm) since last turning due to wear	Continuous	-0.488	-6.60	5.30
kst	Kilometres since last turning	Continuous	109.889	0	248.587
M	Measurement Type (Manual, Laser or Turning)	Nominal	-	-	-
U	Unit Number (1 up to 36)	Nominal	-	-	-
W	Wheelset Position (1 up to 16)	Nominal	-	-	-
S	Side of the wheel (Left or Right)	Nominal	-	-	-
H	Wheels Renewal Cycle: cycle H1 or cycle H2 (more recent)	Nominal	-	-	-
T	Technician (24 different technicians	Nominal	-	-	-

Because the response variables (change in flange height, in flange thickness, in flange slope) and the independent variable kst were calculated based on the difference between the actual measurement and the last post-turning one, the analysis was also dependent on having the complete turning database, which was not the case. Date ranges for measurements did not coincide, with the following ranges being available: manual measurements ranging from July 2001 to June 2016, laser measurements from February 2017 to December 2018 and turning measurements from January 2001 to June 2015. Manual measurements stopped being performed when the laser device was acquired, but turning measurements did not stop, although it was not

possible to make them all available for the current analysis. Hence, this is taken as one of the limitation of the present study. To work around this, some approximations were considered. For the kilometers since turning (kst), it was decided to manually shift the values by bringing them closer to the 0 kst mark (i.e. they were all being calculated with respect to the last turning observation available from 2015, leading to a notably high, and incorrect, value). For the other statistics (ΔF_t, ΔF_h and Δq_R), an average of the values used in different configurations of the turning machine were considered. For instance, the 4 different programs for flange thickness allow the operator to choose from the following post-turning measurements: 31.5mm, 28.5mm, 30.5mm and 29.5 mm. In the absence of a better estimate, the average of those was considered as the post-turning standard to calculate all laser statistics (i.e., 30mm). Same reasoning was applied to ΔF_h and Δq_R. Although these approximations introduce bias to the final results (in comparison to knowing precisely the post-turning measurement), this was the only way available to work around this issue, although the authors plan to revisit the analysis once more turning data is available. Finally, a last correction had to be made to consider a subset of the data: only points with kst < 200,000 km were used. This was necessary because the uncertainty on the actual kst values would increase with more kilometers and, hence, 200,000 seemed like a good trade-off as, in general, all turning interventions are performed before the 200,000 km mark.

Under these corrections, the following exploratory analysis is performed. Figure 58 shows the measurements' histograms for ΔF_t, ΔF_h and Δq_R. Although the laser category has clearly less data than the others (and it was negatively impacted by the lack of post-turning measurements), it still seems to behave better than the other two in terms of range values.

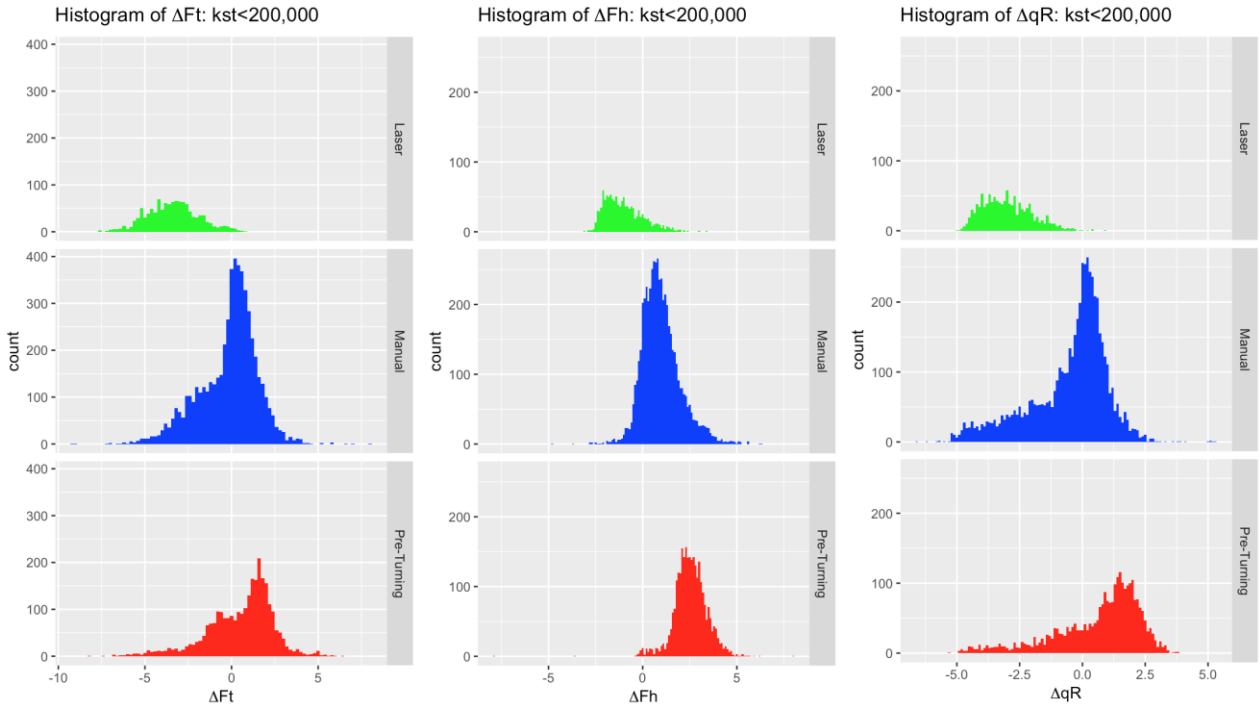


Figure 58: ΔF_h, ΔF_t and Δq_R histograms per type of measurement (kst < 200,000).

Next, measurements boxplots across different units (Figures 59-61) are presented. A quick inspection reveals that unit-to-unit variance is likely to exist and should be considered in the model.

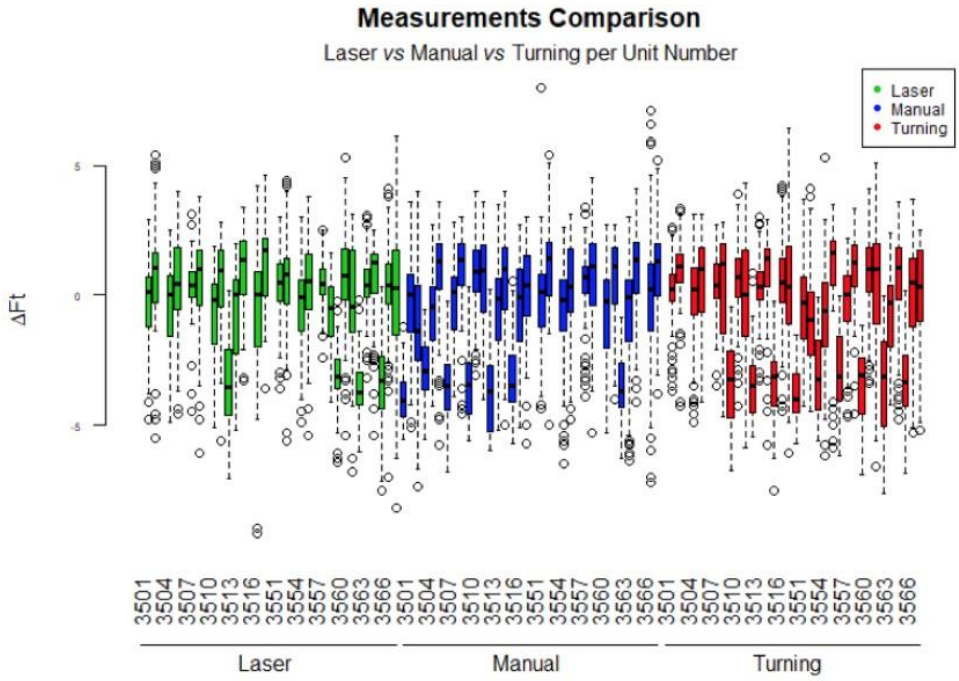


Figure 59: ΔF_t boxplots per type of measurement and unit number (kst < 200,000).

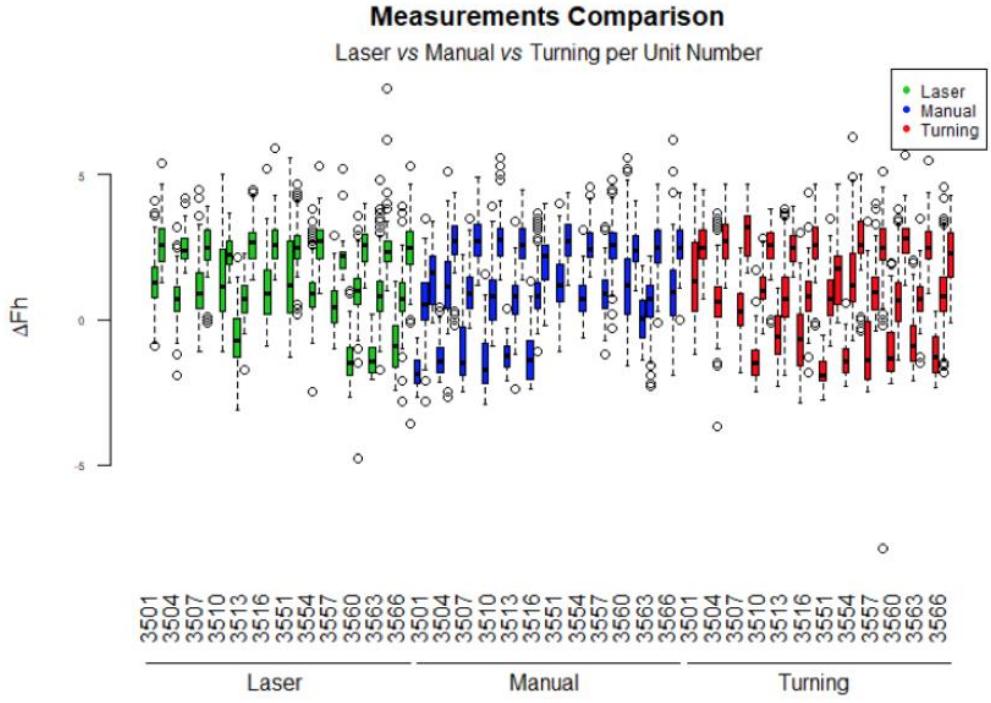


Figure 60: ΔF_h boxplots per type of measurement and unit number (kst < 200,000).

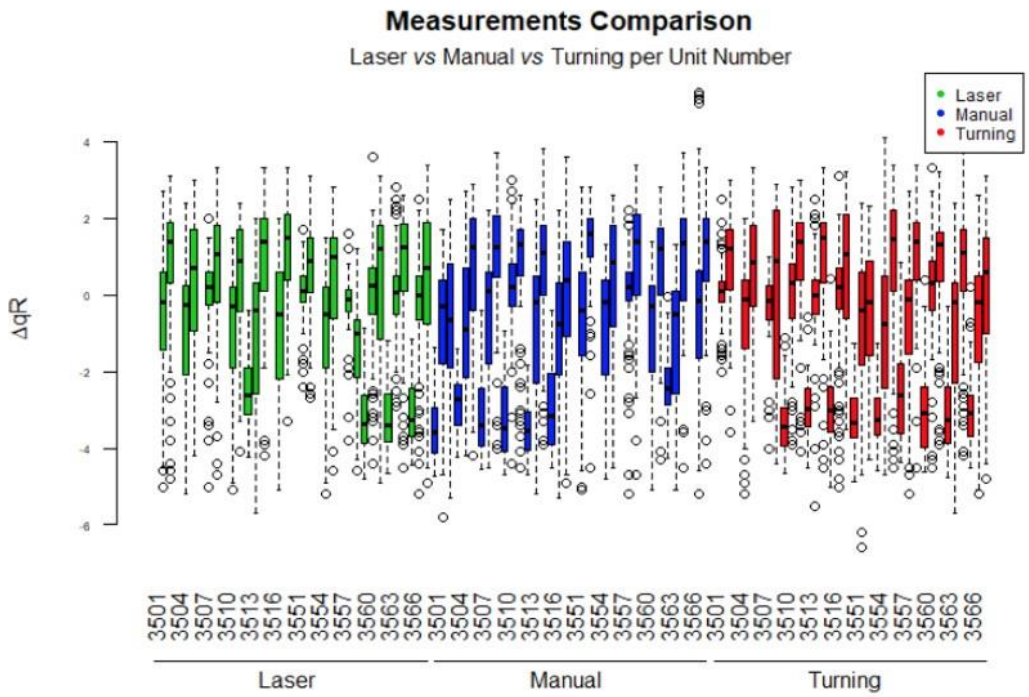


Figure 61: Δq_R boxplots per type of measurement and unit number (kst < 200,000).

Although the exploratory data analysis plays a substantial role in providing a better understanding of the data, it does not allow for inferences regarding the predominance of one type of measurement over the others in terms of precision. Hence, a more systematic approach is considered next.

- **Statistical modelling**

One of the limitations of this study, as mentioned earlier, is that measurements were made in non-overlapping time intervals and, thus, assessing the agreement between measurements within the three methods is only possible if the different conditions of the wheelsets in each timeframe are taken into account. In this study, the latter is done by considering a Linear Mixed Models (LMM) approach having kilometres since last turning (kst) as one of the explanatory variables along with some other relevant categorical variables presented on Table 22, namely: Unit Number (U), Wheelset Position (W), Side of Wheel (S), Wheel's Renewal Cycle (H) and Technician (T).

The option for LMM comes from the fact that they are a flexible, albeit simple method to account for known differences in the variance of the different groups on a linear model. By treating a variable's effect as random (opposed to fixed), the interest shifts from knowing the performance of the various levels of that variable presented on the dataset (think about the performance of different Technicians), to knowing the variation among all levels present in the population. A more intuitive example would be to consider the Technicians in the dataset simply as operators randomly selected from a population. Hence, instead of different "slopes" for each one of the Technicians in the data (fixed effects), the random effects approach allows for random intercepts (i.e., different baseline response values) for each Technician, where the amount of variation in the average response

caused by “Technician” is estimated (random effect) and added to the curve. Further details on LMMs are provided in section 3.1.

Therefore, the assumptions for LMM are the same as in any Linear Model: the explanatory variables are linearly related to the response, the errors (for each model) have constant variance and are independent. It’s also commonly assumed that errors are normally distributed (as above). Hence, if the model is doing a good job, no pattern should be seen on the Residuals vs Fitted values plot (indicating homogeneity of error variance) and on the plot of Residuals vs kst (a visible trend could indicate autocorrelation, for instance). Finally, a normal probability plot of residuals can indicate whether the normality assumption holds.

For the purpose of this study, three different approaches can be compared according to some initial hypotheses, as follows:

1. The change in the flange statistic has an overall mean and the type of measurement variable is not significant in explaining the response.
2. The change in the flange statistic has an overall mean, but there is variation on this mean according to the type of measurement.
3. The change in the flange statistic has a different mean across the different types of measurements.

Although the approaches above might look similar, they differ in the way the variable “type of measurement” (M) variable is considered in the model. More specifically, a) assumes that the final model does not have “type of measurement” as one of its explaining variables, b) assumes that there is a random effect of “type of measurement” that must be accounted for in the model, and c) assumes that “type of measurement” must be treated as a fixed effect in the linear model. In all cases, however, the choice of approach is the LMM, as Unit Number and Technician have both significant random effects that must be accounted for.

Table 23 summarizes the comparison of the different approaches listed above (a-c) for each statistic. The best model is chosen according to the -2 Restricted Log Likelihood (REML) criterion, which can be understood as a “goodness-of-fit” criterion suitable for models including random effects. The final selection comes from an exhaustive search, meaning that all possible combinations of models considering the variables listed on Table 1 were fitted for the specifications in approaches a to c. Only the model with the best performance (in terms of REML) for each approach and each statistic is displayed on Table 23. In general, all final models included the fixed effect of the categorical variables Wheelset Position (W) and Wheel’s Renewal Cycle (H) and the cubic polynomial on kst (kst, kst2, kst3), as well as an intercept. The only exception was the model for Δq_R in approaches b and c, which also included the fixed effect of the categorical variable Side of Wheel (S). In terms of random effects, including the ones for Unit Number (U) and Technician (T) was associated with a better performance of REML in all listed models.

Table 23: Selected models (best REML) for each scenario (a-c) and each statistic.

Dependent Variable	Models	Fixed Effects	Random Effects	REML
ΔF_t	M_a	$1, kst, kst^2, kst^3, W, H$	U, T	36.338
	M_b	$1, kst, kst^2, kst^3, W, H$	U, T, M	35.575
	M_c	$1, kst, kst^2, kst^3, W, H, M$	U, T	35.546
ΔF_h	M_a	$1, kst, kst^2, kst^3, W, H$	U, T	25.945
	M_b	$1, kst, kst^2, kst^3, W, H$	U, T, M	23.208
	M_c	$1, kst, kst^2, kst^3, W, H, M$	U, T	23.187
Δq_R	M_a	$1, kst, kst^2, kst^3, W, H$	U, T	34.112
	M_b	$1, kst, kst^2, kst^3, W, H, S$	U, T, M	33.288
	M_c	$1, kst, kst^2, kst^3, W, H, M, S$	U, T	33.269

A quick inspection on Table 23 reveals that, for all statistics, the model associated with the best REML performance was model c, where the variable M (type of Measurement) is treated as a fixed effect in the model, i.e., the change in the flange statistic has a different mean across the different types of measurements. Although the results for the model b (where M is treated as a Random Effect) were very similar (in terms of REML), the analysis will proceed by considering only the winner in terms of REML. A direct consequence of this, as it should be noted by the reader, is that the initial hypothesis that the distributions of the measurements' types were centred at the same target value will not be further worked on. The working hypothesis becomes that the targets are different (although they may still be close to each other). However, there is a lot of uncertainty on this hypothesis, since (as it will be shown in the residuals plots in the next sections), despite the means being pointed out as different in the final model (i.e., their fixed effects are statistically significant at 5% significance level), the relative spread (or difference) of the means for each type of measurement vs the residuals for each group is not sufficiently large. In other words, the residuals plots do not show the means located at very different points with residuals very close to the mean values for each group (which would indicate that the means were different). Instead, the pattern seen is: mean values close to each other with large dispersion (or residuals spread), which makes the conclusion of different means worth to be further investigated. The latter analysis will be revisited when turning data is completely available. In any case, the interest here still lies on the study of the dispersion around the target values. Thus, for now, the following sections describe the results for the best models in more detail.

- Change in Flange Thickness due to wear (ΔF_t)

The first dependent variable of interest is the change in flange thickness due to wear, ΔF_t. Table 24 provides the estimates for the parameters of the final model M_c.

Table 24: Estimates for the parameters of (best REML) model for ΔF_t

Dependent Variable	Fixed Effects		Random Effects	
ΔF_t	1	$-2.7325(0.1486)$	U	0.255^c
	kst	$42.5896(2.1777)^a$	T	0.438^c
	kst^2	$-4.8998(1.7124)^a$	Residual	1.580^c
	kst^3	$-14.3070(1.7790)^a$	-	-
	W_L	0^b	-	-
	W_M	$0.2113(0.0526)^a$	-	-
	W_T	$0.5667(0.0512)^a$	-	-
	H_1	0^b	-	-
	H_2	$-1.4225(0.0692)^a$	-	-
	M_{laser}	0^b	-	-
	M_{manual}	$2.7548(0.0982)^a$	-	-
	$M_{turning}$	$2.8534(0.1053)^a$	-	-

[a] Approximate standard errors for fixed effects are included in parentheses.
[b] This parameter is considered redundant
[c] Standard Deviation of the estimated effect's variance

As mentioned earlier, the variability in the response ΔF_t is explored through a LMM approach, treating the type of measurement (M) variable as fixed. All fixed effects estimates displayed on Table 24 above are significant at a 5% significance level and the REML linked to this model is 35,546. The random effect associated with Technician accounts for about 7% of total variance, whereas the one associated with Unit Number accounts only for about 2.4%. Hence, measurement noise is still a big component of the model's total variance. It is worth looking at the residuals per measurement type, which are presented on Figure 62.

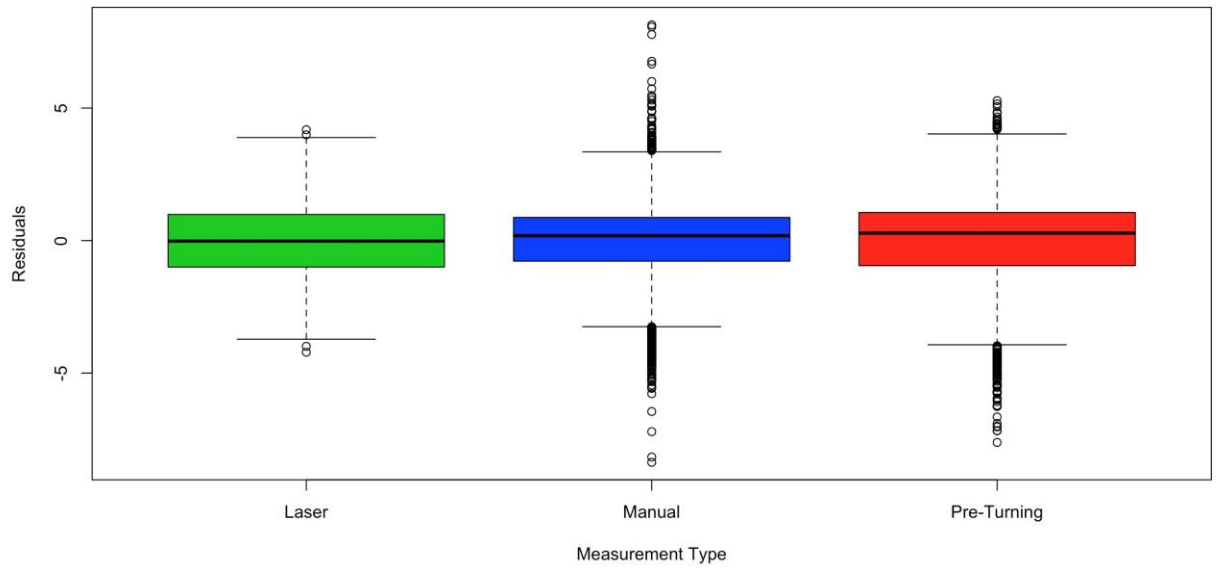


Figure 62: ΔF_t - Analysis of Residuals per Measurement Type.

The difference between the fitted means is clearly small compared to the spread of residuals. This confirms that much of the variability in the data is not explained solely by the types of

measurements. It's interesting, thus, to look at the residuals per categorical variables (different models), which is displayed on Figure 63.

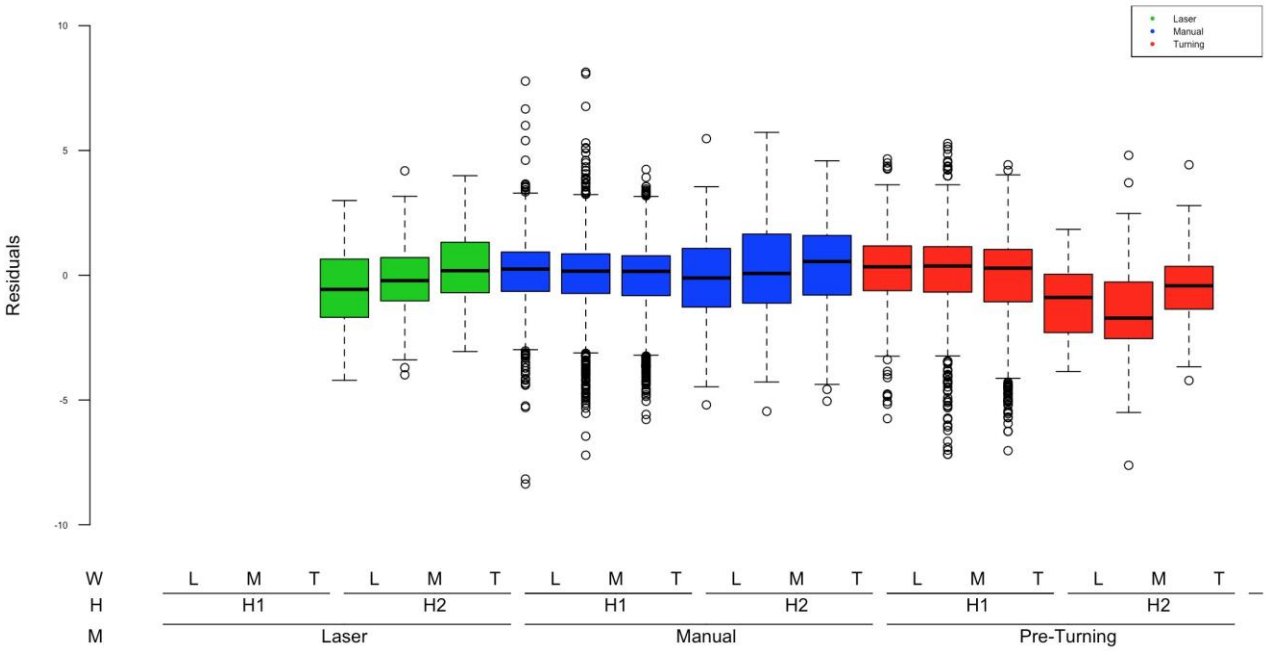


Figure 63: ΔF_t - Analysis of Residuals per Model.

Figure 63 lacks the 3 first boxplots for Laser measurements because all laser measurements were made after the renewal took place, i.e. they were all made under the category "H2" of renewal cycle. Indeed, as Figure 63 suggests, variability was higher among measurements from cycle "H1". It's worth investigating this further in the future. Other than this, the above plot does not add too much information (in comparison to the former) and it will be omitted for the other responses.

Figures 62 and 63 also reveal that, although the interquartile ranges for the residuals of different measurement types are similar, outliers are more frequent in manual and turning measurements. Since those measurements are highly dependent on the Technician's work (at least intuitively), and the random effect associated with Technician was significant, it's interesting to display the residuals of those stratified by Technician as well, as shown on Figure 64.

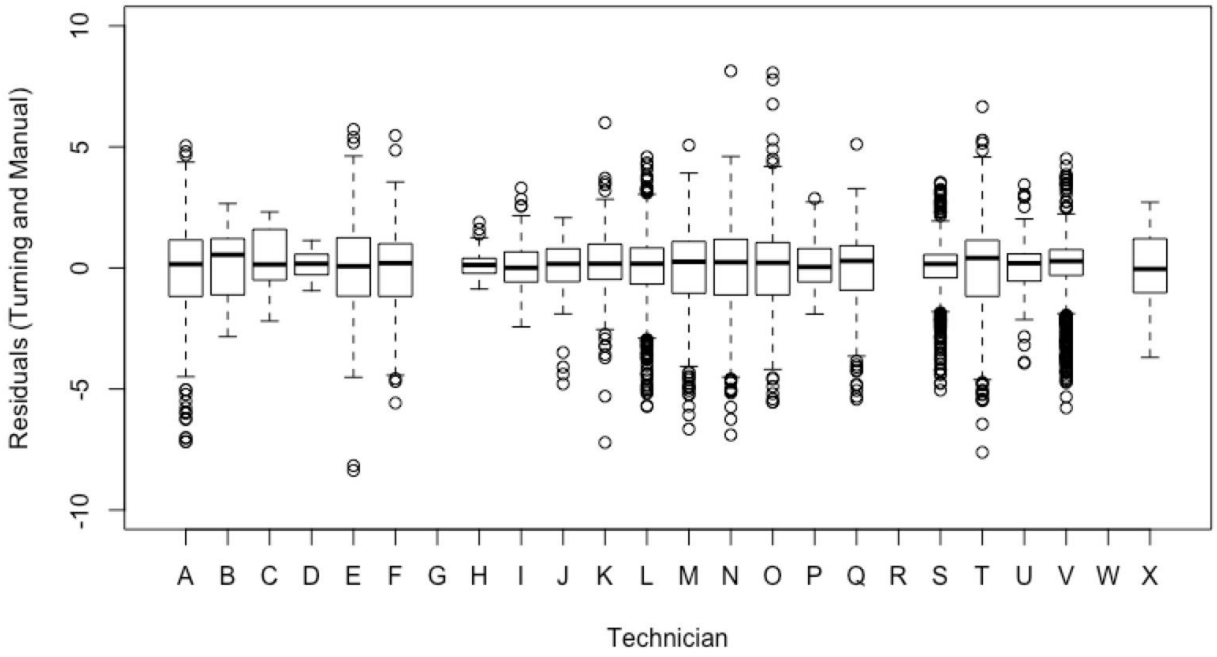


Figure 64: ΔF_t - Analysis of Manual and Turning Residuals per Technician.

When the underlying model is assumed to hold, there is a relevant effect of the Technicians' work. Looking at Figure 64, though manual and turning measurements have a much larger number of observations, there is also a large presence of outliers, which is not desired as these could possibly trigger false maintenance alarms. Laser measurements, on the other hand, are not impacted by the Technicians' work and the device precision is very high. Thus, although impacted by the fewer number of observations in the database, laser measurements seem less likely to produce outliers and (possible) false maintenance alarms.

The same analysis can be repeated for the change in flange height due to wear, ΔF_h. Table 25 provides the estimates for the parameters of the final model M_c. Again, all fixed effects displayed are significant at a 5% significance level and the REML linked to this model is 23.187.

Table 25: Estimates for the parameters of (best REML) model for ΔF_h

Dependent Variable	Fixed Effects		Random Effects	
ΔF_h	1	$-0.6707(0.1147)$	U	0.174^c
	kst	$52.6074(1.1345)^a$	T	0.455^c
	kst^2	$-5.2582(0.8890)^a$	Residual	0.818^c
	kst^3	$7.1793(0.9257)^a$	-	-
	W_L	0^b	-	-
	W_M	$-0.0875(0.0273)^a$	-	-
	W_T	$-0.4655(0.0265)^a$	-	-
	H_1	0^b	-	-
	H_2	$-0.2070(0.0362)^a$	-	-
	M_{laser}	0^b	-	-
	M_{manual}	$2.2408(0.0516)^a$	-	-
	$M_{turning}$	$3.1128(0.0553)^a$	-	-

[a] Approximate standard errors for fixed effects are included in parentheses.
[b] This parameter is considered redundant
[c] Standard Deviation of the estimated effect's variance

For ΔF_h, the random effect associated with Unit Number accounts only for about 3.3%, although here Technician's random effect accounts for about 22.8% of total variance, very different from what was seen on ΔF_t analysis. Still, measurement noise is a big component of the model's total variance. The residuals per measurement type are presented on Figure 65.

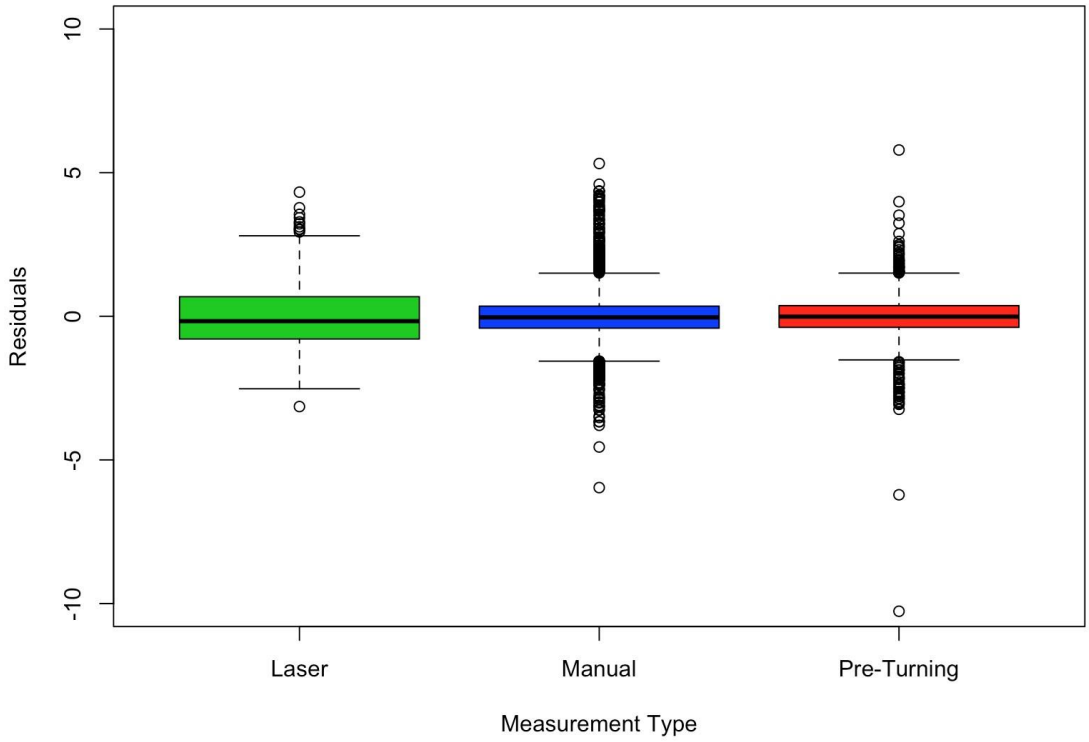


Figure 65: ΔF_h - Analysis of Residuals per Measurement Type.

Although Laser residuals present a bigger interquartile range, the pattern of outliers in Manual and Turning measurements still seems to be present. A comparison of the residuals for manual and turning measurements per Technician is presented on Figure 66.

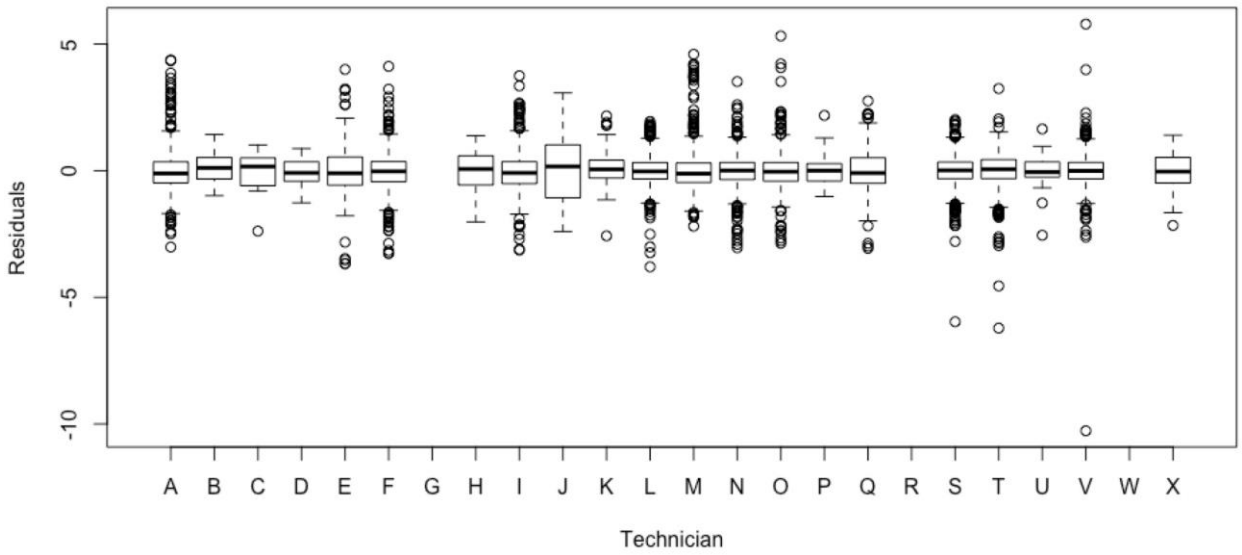


Figure 66: ΔF_h - Analysis of Manual and Turning Residuals per Technician.

When the underlying model is assumed to hold, the random effects of Technicians are even more significant than those found for the ΔF_t analysis. The conclusions are similar: although manual and turning measurements have a much larger number of observations, there is also a large presence of outliers, negatively impacting the precision of those measurements. Laser measurements seem to be less susceptible to the presence of outliers, and hence, more robust against false maintenance alarms.

The last analysis is on the change in Flange Slope due to wear, Δq_R. Table 26 provides the estimates for the parameters of the final model M_c, this time with the addition of the categorical variable “Side of Wheels” (S), which was shown to improve REML. Again, all fixed effects displayed are significant at a 5% significance level and the REML linked to this model is 33.269.

Table 26: Estimates for the parameters of (best REML) model for Δq_R

Dependent Variable	Fixed Effects		Random Effects	
Δq_R	1	$-2.5890(0.1312)$	U	0.234^c
	kst	$51.3865(1.9287)^a$	T	0.376^c
	kst^2	$-5.1604(1.5168)^a$	Residual	1.399^c
	kst^3	$-7.0800(1.5757)^a$	-	-
	W_L	0^b	-	-
	W_M	$0.1036(0.0466)^a$	-	-
	W_T	$0.4610(0.0454)^a$	-	-
	H_1	0^b	-	-
	H_2	$-1.0558(0.0613)^a$	-	-
	S_L	0^b	-	-
	S_R	$-0.0662(0.0288)^a$	-	-
	M_{laser}	0^b	-	-
	M_{manual}	$2.2800(0.0869)^a$	-	-
	$M_{turning}$	$2.7721(0.0932)^a$	-	-

[a] Approximate standard errors for fixed effects are included in parentheses.
[b] This parameter is considered redundant
[c] Standard Deviation of the estimated effect's variance

For Δq_R, the random effect associated with Unit Number accounts only for about 2.5% and for Technician only about 6.5% of total variance, somewhat like ΔF_t analysis. The residuals per measurement type are presented on Figure 67.

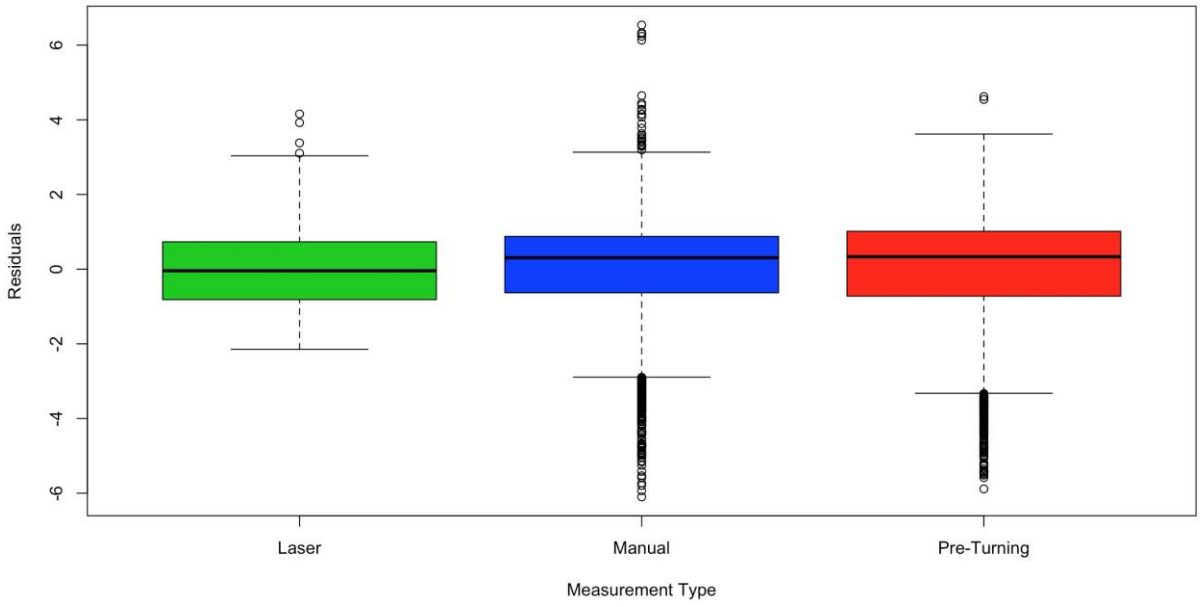


Figure 67: Δq_R - Analysis of Residuals per Measurement Type.

And the corresponding comparison of manual and turning residuals per technician is presented on Figure 68:

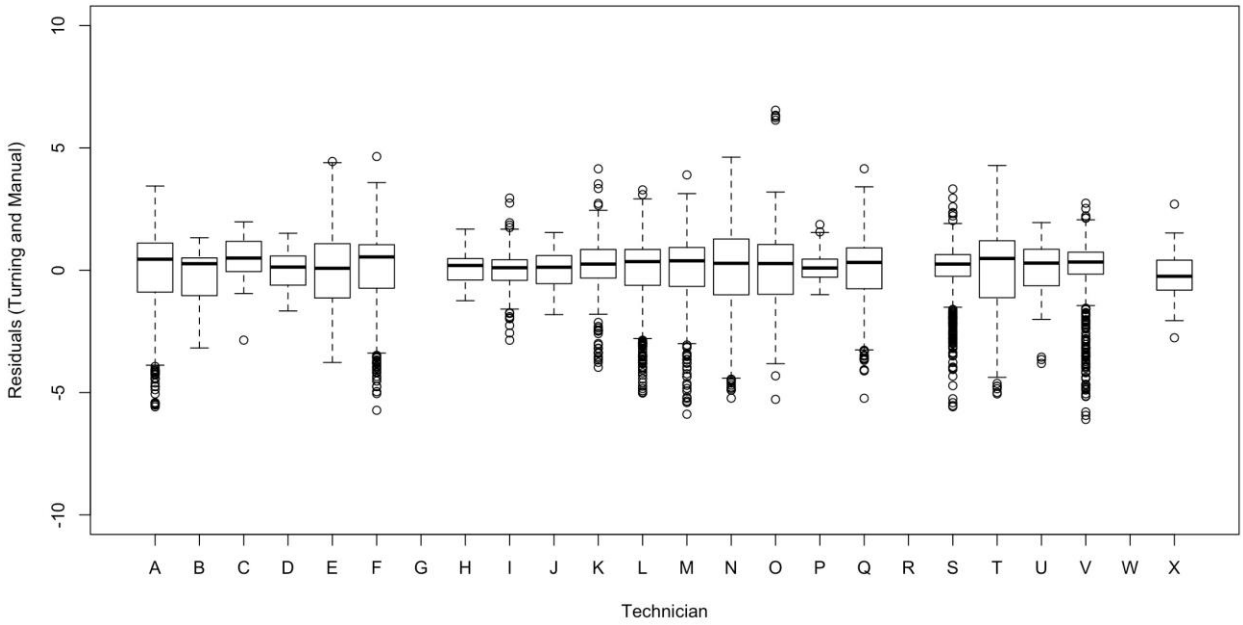


Figure 68: Δq_R - Analysis of Manual and Turning Residuals per Technician.

Again, assuming that the underlying model is true, the conclusions are very similar to the ones made for ΔF_t and ΔF_h. Laser measurements seem to be less susceptible to the presence of outliers, and hence, more robust against false alarms for maintenance.

This section presented a systematic approach, using Linear Mixed Models, to compare the variances of types of measurement in the context of railway wheelsets maintenance strategies and in terms of three main statistics (responses): ΔF_t, ΔF_h and Δq_R. The goal was to be able to make inferences about the quality of measurements as a matter of the dispersion around the mean/target values. Some interesting findings are related to the best model (in terms of REML criterion) being the one treating the variable "type of measurement" as a fixed effect, meaning that the best models (for all three responses) were the ones considering different slopes based on measurement types and some other categorical variables, which was not intuitively expected. The next relevant result showed that the behavior of the residuals (assuming the underlying models are correct) did not exhibit the expected pattern, i.e., they did not allow for inference that the measurements differ significantly from one type to another, despite means being different according to the selected model. However, another important result came from the observation of residual plots stratified by Type of Measurements: the presence of outliers was more common in the manual and turning cases, where measurements are dependent on the work of Technicians. Moreover, the random effect associated with Technician was significant in all models, and plotting the residuals by Technician revealed that part of the dispersion can be explained by the human work and that there are notable differences from one Technician to the others.

However, one of the limitations of this study, as mentioned several times, was the lack of turning measurements to compute the actual $kst, \Delta F_t$, ΔF_h and Δq_R for the laser measurements. The authors have made several approximations, but ultimately, when turning data is available, it is worth

revisiting the analysis, as to confirm the results obtained and further investigate the outcomes that were not initially expected.

4. CONCLUSIONS

In the present report, a major division in Condition-Based Maintenance (CBM) model application was made between Part I on prognosis using on-train diagnostic data and Part II on decision support based on condition data. Several statistical techniques and decision support tools within the CBM model were applied to different subsystems or components of the vehicle system.

In part I, an overall discussion on data collection and formatting was provided, becoming evident that in order to apply the proposed CBM framework and techniques a strong and robust information system has to be available within railway organisations. A discussion on failure/event of interest is provided, namely on the selection, characterisation and associated prognostic model for case study from a regional EMU. This suggests that CBM implementation requires databases with comprehensive classification of event, failures and resulting maintenance actions to take full advantage of the proposed techniques. Moreover, by analysing a case study on high speed EMU, relevant CBM techniques and frameworks were explored such as: the diagnostic data architecture, data science methods with engineering design and other machine learning techniques. This also showed that CBM full implementation requires a comprehensive diagnostic architecture, rich enough to take advantage of the proposed techniques.

In part II, a full case study on wheelset maintenance is provided for the Fertagus train operating company. By analysing and focusing on the wheelset component, different techniques were explored to support decision making, namely: statistical modelling of wheelset wear for the Fertagus case study and for the London Underground case study, survival modelling of wheelset damage leading to a Markov Decision Process (MDP) approach to derive an optimal map of a maintenance turning strategy for the wheelset. Such predictive techniques show that obtaining reliable inspection or condition data are crucial to support full application of such techniques. Furthermore, supporting maintenance decisions within a tactical maintenance plan and operational maintenance scheduling are comprehensive activities that require that maintenance and inspection records are digitalized to facilitate an easier implementation of CBM model within railway organisations. Finally, reliable inspection devices are crucial to limit uncertainty associated with measuring condition data, suggesting that CBM implementation requires reliable sensors (e.g. laser equipment for the wheelset condition and profile measurements).

The following sections explore CBM implementation (section 4.1) and finalises with main conclusions and further research (section 4.2).

4.1 TOWARDS CBM IMPLEMENTATION

It is crucial to discuss how the CBM techniques and proposed framework can be implemented for other case studies and in other railway organisations. By applying the case studies, explored in Parts I and II, such techniques and main lessons can be highlighted (in subsection 4.1.1) and a

discussion on guidelines to CBM implementation is also provided (in subsection 4.1.2). Finally, barriers to CBM implementation are identified in subsection 4.1.3.

4.1.1 LESSONS FROM CASE STUDIES

Several lessons can be taken from the application of the CBM techniques and models to the cases studies on diagnostic data (explored in Part I) and condition data (explored in Part II). A first important lesson is that the CBM techniques and framework revealed to be flexible enough to be applied to diagnostic and condition data, and to different components and subsystems within the railway vehicle system. This flexible nature of the CBM techniques suggests that the future application to other systems and other components would not be difficult, with the natural adaptations specific to each of these systems.

For instance, statistical models such as Linear Mixed Models (LMMs) and Survival Models (SM) are very flexible to take into account and test the significance of any controlling variable that the modeller and/or practitioner thinks that might influence the life-cycle or the deterioration of a given system. Moreover, the Markov Decision Process (MDP) provides a robust technique to derive an optimal maintenance strategy map (i.e. showing which decision should be made depending on the exact condition of that system). For the wheelset turning strategy, this optimal maintenance map provides a strategy, based on the kilometres since last turning and on the diameter of the wheelset, supporting maintenance technicians to make turning decisions from a life-cycle perspective and based on the inspection records and observed trends in wear and damage. It should be mentioned that the Transitions Matrices were estimated using statistical techniques and survival statistical modelling to estimate hazard functions and associated transition probabilities. Therefore, the application of MDP for other components might require a more extensive knowledge of main failure modes and main quality and performance indicators specific to that component.

Regarding maintenance planning models, it was showed that such Integer Linear Programming (ILP) formulations can easily integrate the complexity of maintenance planning in a medium-term horizon, adding technical constraints associated with depot and its configuration, human resources, amount of work, etc. These comprehensive models were detailed for the Fertagus case study, and it would not be difficult to adapt them to other case studies/railway companies. It was shown that the shunting costs represent a major fraction of the costs. Moreover, and regarding the operational maintenance scheduling models, it was shown that the assignment of train units to normal services and maintenance tasks can provide additional flexibility to overcome potential delays or uncertainty associated with maintenance durations. These two maintenance decision support models were applied to the Fertagus case study: i) a tactical maintenance planning model to find an optimal maintenance plan and ii) an operational maintenance scheduling model to assign train units to daily services and maintenance tasks to find an optimal schedule. Both models support the assessment of different maintenance intervals and inspection intervals in the overall maintenance plan, and they can support the impact assessment in Deliverable D4.2.

The benefits of correlating the diagnostic event data with failure records in the machine learning approaches (such as a Recurrent Neural Network) was demonstrated in Part I. In this supervised method, information on failures that result in maintenance are used as a resource to teach the model the sequence of events or patterns leading up to a failure. Using this information, it is possible to provide an earlier warning of an imminent failure to allow maintenance plans to be adjusted accordingly.

4.1.2 GUIDELINES TO IMPLEMENTATION

Shifting from current maintenance practices to Condition-Based Maintenance (CBM) strategies is not an easy task for railway organisations. In fact, the full application of these CBM techniques may require specific training to different human resources, which play different roles in the inspection and maintenance processes (supervisors, technicians, planners, etc.). In particular, it is crucial to facilitate interpretation of optimal decision maps and to ensure the correct application of several statistical techniques.

Statistical techniques can only be applied if there are rich and reliable datasets within the organisation with inspection, diagnostic, failure, maintenance and renewal records. Digitalisation of such records is fundamental to allow an easier implementation. Sometimes, some calibrations or interpretation is needed and some sort of feedback is required from maintenance technicians, supervisors or planners, to take full advantage from the records, and potentially remove unreliable part of the dataset.

Along with the reliability, the accessibility of the data acquired from the different inspection or on-train systems is also an important factor. If the data is not made available to the various assessment methods in the required time frame the potential benefits, in terms of provided enough time and information to support maintenance planning, will be reduced. Ideally, the assessment methods would be granted periodic access to the data through a central server/database or a cloud-based system.

4.1.3 BARRIERS

The full application of these techniques may require specific training to different human resources, which play different roles in the inspection and maintenance processes (supervisors, technicians, planners, etc.). In particular, it is crucial to facilitate interpretation of optimal decision maps and to ensure the correct application of several statistical techniques. Moreover, skills in data science may also be an advantage to speed up the process of implementing such CBM model in railway organizations.

One important limitation towards the full implementation of the CBM techniques and framework is the digitalisation of inspection, maintenance and failure records and the availability of long series

of historic data, both for diagnostic and/or condition. On the one hand, machine learning and data-driven approaches require training subsets and long series of past records can provide that. On the other hand, such records provide a window of observation and validity for predictions. Outside this window of observation, such predictive approaches may fail, and thus, decision supporting is also only valid in this observation window.

It should also be mentioned that specific characteristics of the fleets analysed and associated case studies (e.g. size, depot configuration, services) may limit the straightforward CBM implementation of such techniques to other train fleets, and thus, some sort of adaptation might be required. It is also important to mention that larger sizes of the fleet may limit the performance of the integer linear programming techniques used in the decision supporting techniques.

In relation to the application of on-train diagnostic data in the CBM model; whilst most modern rolling stock will be equipped with a diagnosis system, which collects faults, events and operational data, the functionality of these systems may vary considerably (e.g. safety critical systems, environment variables). Therefore, the level of detail collected by one fleet of trains may not be different for other fleets and therefore the assessment methods might need to be tailored to specific fleets.

4.2 MAIN CONCLUSIONS AND FURTHER RESEARCH

The present report explored applications of the CBM model and its techniques to support train operating companies to shift their current maintenance strategies to condition-based maintenance strategies. It aimed to apply the CBM techniques and framework proposed in Deliverable D2.2 for several case studies. Such application showed some success, demonstrating the wide variety systems and components (within the railway vehicle systems) in which it can be applied.

Part I explored the use of machine learning techniques to predict sequences in on-train diagnostic data which could provide an indication of an imminent component/system failure. When using the techniques in a supervised method, where the sequence of diagnostics events are correlated to maintenance actions, the model is able to provide an early indication of an imminent failure based on the training dataset used in the project. Whist the unsupervised method was good in extracting trends and patterns from the diagnostics data, further interpretation of the outputs is required in order to link these patterns to specific failures and resulting maintenance actions. The case studies conducted in Part I also highlighted the importance in the quality and volume of data available for input in to the model.

Part II explored decision support based on condition data, with the specific case study on wheelset maintenance. By using Linear Mixed Models, it was possible to estimate the evolution of the different geometrical indicators, by controlling for the effect of different explaining variables. A comparison with London Underground inspection wheelset data suggested that such models can be easily adapted to different fleets. It provided a more comprehensive understanding on the topic of exploring wear trajectories of railway wheelsets, finding statistical patterns consistent with other

fleets. A survival analysis estimated the survival probabilities associated with damage occurrence as the kilometres since last turning/renewal increases. A Markov Decision Process (MDP) approach provided an optimal decision map to trigger preventive turning depending on the current wheelset diameter and the kilometres since last turning. Two maintenance decision support models were applied to the Fertagus case study: i) a tactical maintenance planning model to find an optimal maintenance plan and ii) an operational maintenance scheduling model to assign train units to daily services and maintenance tasks to find an optimal schedule. Both models support the assessment of different maintenance intervals and inspection intervals in the overall maintenance plan, and they can support the impact assessment in Deliverable D4.2. Finally, an assessment of uncertainty associated with different inspection techniques was conducted, showing that laser inspection provides a lower number of outliers and thus more reliability in wheelset inspection.

It is very important to mention that the present research stream targeted TRL level 2-3 for the application of the CBM model, with a proof of concept and feasibility of such models within the normal practice and procedures in railway organisations such as the Fertagus train operating company. Further research needs to target higher TRL levels by integrating such techniques in a more robust way.

5. REFERENCES

Andrade, A. R., and Stow, J., 2016. Statistical Modelling of Wear and Damage Trajectories of Railway Wheelsets. Qual. Reliab. Eng. Int. 32(8), pp. 2909–2923.

Andrade, A. R., and Stow, J., 2017a. Assessing the Potential Cost Savings of Introducing the Maintenance Option of 'Economic Tyre Turning' in Great Britain Railway Wheelsets. Reliab. Eng. Syst. Saf. 168, pp. 317–325.

Andrade, A. R., and Stow, J., 2017b. Assessing the Efficiency of Maintenance Operators: A Case Study of Turning Railway Wheelsets on an under-Floor Wheel Lathe. Proc. Inst. Mech. Eng. Part O J. Risk Reliab. 231(2), pp. 155–163.

Asplund, M., Palo, M., Famurewa, S., and Rantatalo, M., 2016. A Study of Railway Wheel Profile Parameters Used as Indicators of an Increased Risk of Wheel Defects. Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit 230(2), pp. 323–334.

Bates, D., 2010. Lme4: Mixed-Effects Modelling With R [Online]. Available: http://lme4.r-forge.r-project.org/book. [Accessed: 31-Jul-2018].

Bates, D., 2018. Computational Methods for Mixed Models pp. 1–21 [Online]. Available: https://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf. [Accessed: 31-Jul-2018].

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting Linear Mixed-effects Models Using lme4. arXiv preprint 2014 arXiv:1406.5823.

Braga, J. A. P., and Andrade, A. R., 2019. Optimizing Maintenance Decisions in Railway Wheelsets: A Markov Decision Process Approach. Proc. Inst. Mech. Eng. Part O J. Risk Reliab., 233(2) pp. 285-300.

Braghin, F., Lewis, R., Dwyer-Joyce, R. S., and Bruni, S., 2006. A Mathematical Model to Predict Railway Wheel Profile Evolution Due to Wear. Wear 261(11–12), pp. 1253–1264.

BS-EN15313:2012, 2012. Testing for the Acceptance of Running Characteristics of Railway Vehicles – Testing of Running Behaviour and Stationary Tests.

Chadès, I., Chapron, G., Cros, M., Garcia, F. and Sabbadin, R. (2014). MDPtoolbox: a multi-platform toolbox to solve stochastic dynamic programming problems. Ecography. 37, pp. 916–920.

Cox, D. R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society. Series B (Methodological). 34(2), pp. 187-220.

Cox, D. R. and Oakes, D. (1984). Analysis of Survival Data. Chapman & Hall, London.

Ferreira, J. C., Freitas, M. A., and Colosimo, E. A., 2012. Degradation Data Analysis for Samples under Unequal Operating Conditions: A Case Study on Train Wheels. J. Appl. Stat., 39(12) pp. 2721–2739.

Freitas, M., Toledo, M., Colosimo, E., and Pires, M., 2009. Design and Analysis for the Gaussian Process Model. Qual. Reliab. Eng. Int. 25, pp. 607–629.

Galecki, A., and Burzykowski, T., 2013. Linear Mixed-Effects Models Using R, Springer, New York.

Hossein Nia, S., Casanueva, C., and Stichel, S., 2015. Prediction of RCF and Wear Evolution of Iron-Ore Locomotive Wheels. Wear 338–339, pp. 62–72.

Iwnicki, S., 2006. Handbook of Railway Vehicle Dynamics.

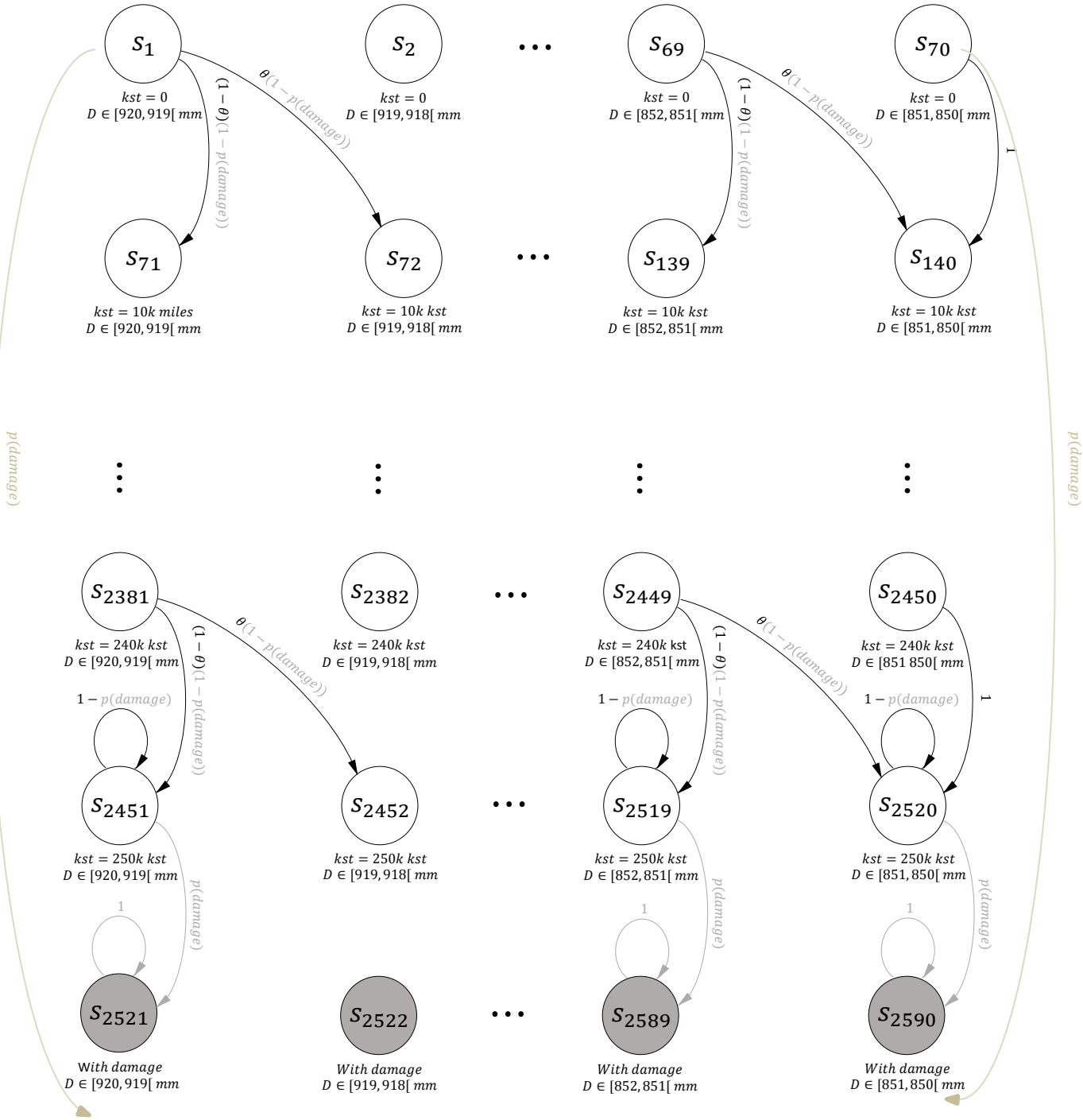
Jiang, Z., Banjevic, D., Mingcheng, E. M., and Li, B., 2017. Optimizing the Re-Profiling Policy Regarding Metropolitan Train Wheels Based on a Semi-Markov Decision Process. Proc. Inst. Mech. Eng. Part O J. Risk Reliab. 231(5), pp. 495–507.

Lewis, R., and Olofsson, U., 2004. Mapping rail wear regimes and transitions. Wear. 257(7-8), pp. 721-729.
Lin, J., Asplunda, M., and Parida, A., 2014. Reliability Analysis for Degradation of Locomotive Wheels Using Parametric Bayesian Approach. Qual. Reliab. Eng. Int. 30(5), pp. 657–667.
Mingcheng, E., Li, B., Jiang, Z., Li, Q., 2018. An Optimal Reprofiling Policy for High-Speed Train Wheels Subject to Wear and External Shocks Using a Semi-Markov Decision Process. IEEE Transactions on Reliability. 67(4), pp. 1468-1481.
Molyneux-Berry, P., and Bevan, A., 2012. Wheel Surface Damage: Relating the Position and Angle of Forces to the Observed Damage Patterns. Veh. Syst. Dyn. 50(SUPPL. 1), pp. 335–347.
Müller, S., Scealy, J. L., Welsh, A., H., 2013. Model selection in linear mixed models. Statistical Science 28(2):135–167.
Pascual, F., and Marcos, J., 2004. Wheel Wear Management on High-Speed Passenger Rail: A Common Playground for Design and Maintenance Engineering in the Talgo Engineering Cycle. ASME/IEEE Joint Rail Conference, Cackovic DL, Ku B-Y (Eds), American Society of Mechanical Engineers, Baltimore pp. 193–200.
Pombo, J., Ambrósio, J., Pereira, M., Lewis, R., Dwyer-Joyce, R., Ariaudo, C., and Kuka, N., 2011a. Development of a Wear Prediction Tool for Steel Railway Wheels Using Three Alternative Wear Functions. Wear 271(1–2), pp. 238–245.
Pombo, J., Ambrósio, J., Pereira, M., Verardi, R., Ariaudo, C., and Kuka, N., 2011b. Influence of Track Conditions and Wheel Wear State on the Loads Imposed on the Infrastructure by Railway Vehicles. Comput. Struct. 89(21–22), pp. 1882–1894.
Schulte-Werning, B.; Mösken, H.; Linke, P.; Hoksch, V.; Uebel, L.; Wolter, K.-U.: On current Developments in Condition-Based Maintenance at DB. Proceedings of 11th World Congress on Railway Research 2016, Paper 450, 29 May-2 June 2016, Milano, Italy.
Tea, C., 2012. REX et Données Subjectives: Quel Système d'information Pour La Gestion Des Risques? Numéro 2012–04 Des Cahiers de La Sécurité Industrielle. Fond. Pour une Cult. Sécurité Ind. Toulouse, Fr. (ISSN 2100-3874) p. 113 [Online]. Available: https://www.foncsi.org/fr/publications/collections/cahiers-securite-industrielle/REX-donnees-subjectives/CSI-REX-donnees-subjectives.pdf. [Accessed: 01-Aug-2018].
Wang, L., Xu, H., Yuan, H., Zhao, W., and Chen, X., 2015. Optimizing the Re-Profiling Strategy of Metro Wheels Based on a Data-Driven Wear Model. Eur. J. Oper. Res. 242(3), pp. 975–986.
Yin, G. and Zhang, Q. (2006). Discrete-Time Markov Chains. New York: Springer.

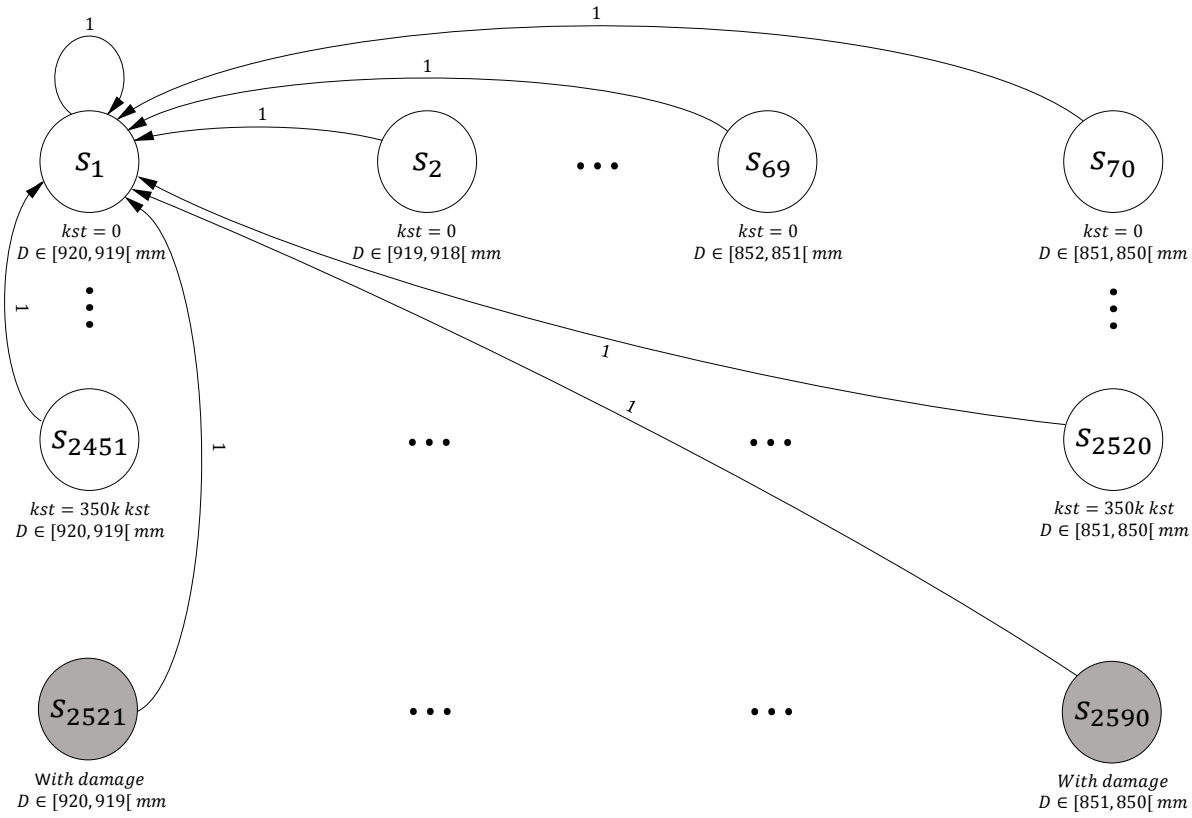
APPENDIX A – State Space division and transition probabilities

A1 State space division and transition probabilities for the 'Do nothing' action ($a = 1$)

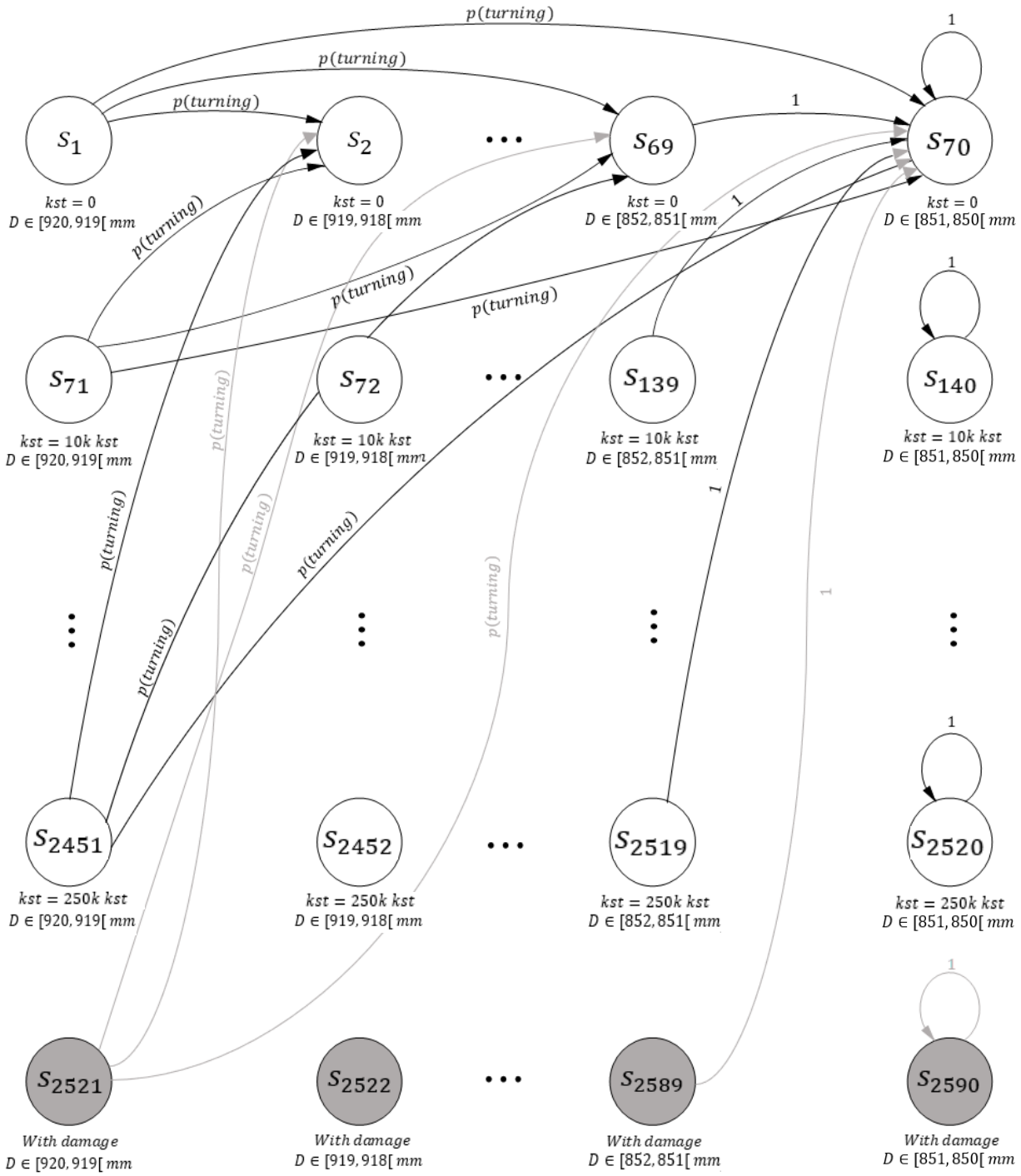
Note: grey colour represents the effect of damage.



A2 State space division and transition probabilities for the 'Renewal' action ($a = 2$)
Note: grey colour represents the effect of damage.



A3 State space division and transition probabilities for the 'Turning' action ($a = 3$)
Note: grey colour represents the effect of damage.



Appendix B – Tables with information on the case study of a Portuguese train operating company.

Table B.1: Information concerning stations, their names number and minimal turning times.

Station Name	Station Number, s	Minimal Turning Time, TM_s (min)
Roma-Areeiro	1	1
Entrecampos	2	1
Sete-Rios	3	1
Campolide	4	1
Pragal	5	1
Corroios	6	1
Foros de Amora	7	1
Fogueteiro	8	1
PMC (depot)	9	1
Coina	10	1
Penalva	11	1
Pinhal-Novo	12	1
Venda do Alcaide	13	1
Palmela	14	1
Setúbal	15	1

Table B.2: Pairs of stations between which there can exist dead-headings.

$W_{s,s'}$		s'														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
s	**1**	0	0	0	0	1	0	0	1	1	1	0	0	0	0	1
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	9	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	10	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0

Table B.3: Associated lengths between pairs of stations.

$CW_{s,s'}$ (km)		s'														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	1	0	1.13	2.84	4.04	11.68	16.78	19.38	22.12	25.6	27.32	32.62	41.42	45.23	49.33	54.16
	2	1.13	0	1.71	2.91	10.55	15.65	18.25	20.99	0	26.19	31.49	40.29	44.1	48.2	53.03
	3	2.84	1.71	0	1.2	8.84	13.95	16.55	19.28	0	24.49	29.79	38.59	42.39	46.49	51.32
	4	4.04	2.91	1.2	0	7.65	12.75	15.35	18.09	0	23.29	28.59	37.39	41.19	45.29	50.12
	5	11.68	10.55	8.84	7.65	0	5.1	7.7	10.44	0	15.64	20.94	29.74	33.54	37.64	42.47
	6	16.78	15.65	13.95	12.75	5.1	0	2.6	5.34	0	10.54	15.84	24.64	28.44	32.54	37.37
	7	19.38	18.25	16.55	15.35	7.7	2.6	0	2.74	0	7.94	13.24	22.04	25.84	29.94	34.77
s	8	22.12	20.99	19.28	18.09	10.44	5.34	2.74	0	0	5.2	10.5	19.3	23.1	27.2	32.03
	9	25.6	0	0	0	0	0	0	0	0	1.7	0	0	0	0	28.6
	10	27.32	26.19	24.49	23.29	15.64	10.54	7.94	5.2	1.7	0	5.3	14.1	17.9	22	26.83
	11	32.62	31.49	29.79	28.59	20.94	15.84	13.24	10.5	0	5.3	0	8.8	12.6	16.7	21.53
	12	41.42	40.29	38.59	37.39	29.74	24.64	22.04	19.3	0	14.1	8.8	0	3.8	7.9	12.73
	13	45.23	44.1	42.39	41.19	33.54	28.44	25.84	23.1	0	17.9	12.6	3.8	0	4.1	8.93
	14	49.33	48.2	46.49	45.29	37.64	32.54	29.94	27.2	0	22	16.7	7.9	4.1	0	4.83
	15	54.16	53.03	51.32	50.12	42.47	37.37	34.77	32.03	28.6	26.83	21.53	12.73	8.93	4.83	0

Table B.4: Associated durations between pairs of stations.

$DW_{s,s'}$ (min)		s'														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	1	0	0	0	0	16	0	0	22	24	26	0	0	0	0	45
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s	8	22	0	0	0	0	0	0	0	0	5	0	0	0	0	21
	9	24	0	0	0	0	0	0	0	0	2	0	0	0	0	21
	10	26	0	0	0	0	0	0	5	2	0	0	0	0	0	17
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	45	0	0	0	0	0	0	21	21	17	0	0	0	0	0

Table B.5: Constants used.

Constant	Unit	Value
NU	-	17
NS	-	15
NT	-	790
ND	day	5
NM	-	14
LN	-	10000
PW	-	850
$PTHOM$	-	300
$PTZM$	-	850

Table B.6: Information about tasks (the complete is provided in the electronic version).

Task (T_i)	DEM_i	CAP_i	Sd_i	Sa_i	Dd_i (min)	Da_i (min)
1	1	2	1	15	343	401
2	2	2	15	1	418	476
3	2	2	1	10	483	516
4	2	2	10	1	523	556
5	2	2	1	10	563	596
6	2	2	10	1	1053	1086
7	2	2	1	15	1093	1151
8	1	2	15	1	1168	1226
9	1	2	1	10	1233	1266
10	1	2	10	1	1283	1316
11	1	2	1	10	1333	1366
12	1	2	15	15	1152	1258
13	1	2	15	1	388	446
14	1	2	1	10	453	486
15	1	2	15	1	508	566
16	1	2	1	10	573	606
17	2	2	10	1	963	996
18	2	2	1	15	1003	1061
19	2	2	15	1	1078	1133
20	2	2	1	10	1143	1176
(...)						
817	0	0	12	12	0	0
818	0	0	13	13	0	0
819	0	0	14	14	0	0
820	0	0	15	15	0	0

Table B.7: Maintenance actions that need to be performed on the planning period.

$KM_{k,m}$		m													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
k	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1
	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table B.8: Duration and working load (amount of work) of each maintenance action.

Maintenance Action, m	MT_m (min)	AW_m (min)
1	150	744
2	420	1680
3	210	840
4	210	840
5	276	840
6	186	744
7	186	744
8	186	744
9	186	744
10	186	744
11	420	840
12	53	210
13	53	210
14	60	60

Appendix C – Output of 1-day fleet assignment from operational scheduling model for Portuguese train operating company