Regression Analysis of Caterpillar 793D Haul Truck Engine Failure Data and Through-Life Diagnostic Information Using the Proportional Hazards Model

by

Wiehahn Alwyn Carstens

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Industrial Engineering at Stellenbosch University

Department of Industrial Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Dr. P.J. Vlok, Prof. C.S.L. Schutte and T. Visser

March 2012

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: ....

Abstract

Regression Analysis of Caterpillar 793D Haul Truck Engine Failure Data and Through-Life Diagnostic Information Using the Proportional Hazards Model

W.A. Carstens

Department of Industrial Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MScEng (Industrial)

March 2012

Physical Asset Management (PAM) is becoming a greater concern for companies in industry today. The widely accepted British Standards Institutes' specification for optimized management of physical assets and infrastructure is PAS55. According to PAS55, PAM is the "systematic and co-ordinated activities and practices through which an organization optimally manages its physical assets, and their associated performance, risks and expenditures over their life cycle for the purpose of achieving its organizational strategic plan".

One key performance area of PAM is Asset Care Plans (ACP). These plans are maintenance strategies which improve or ensure acceptable asset reliability and performance during its useful life. Maintenance strategies such as Condition Based Maintenance (CBM) acts upon Condition Monitoring (CM) data, disregarding the previous failure histories of an asset. Other maintenance strategies, such as Usage Based Maintenance (UBM), is based on previous failure histories, and does not consider CM data.

Regression models make use of both CM data and previous failure histories to develop a model which represents the underlying failure behaviour of the asset under study. These models can be of high value in ACP development due to the fact that Residual Useful Life (RUL) can be estimated and/or the long term life cycle cost can be optimized.

The objective of this thesis was to model historical failure data and CM data well enough so that RUL or optimized preventive maintenance instant estimations can be made. These estimates were used in decision models to develop maintenance schedules, i.e. ACPs.

Several regression models were evaluated to determine the most suitable model to achieve the objectives of this thesis. The model found to be most suitable for this research project was the Proportional Hazards Model (PHM). A comprehensive investigation on the PHM was undertaken focussing on the mathematics and the practical implementation thereof.

Data obtained from the South African mining industry was modelled with the Weibull PHM. It was found that the developed model produced estimates which were accurate representations of reality. These findings provide an exciting basis for the development of future Weibull PHMs that could result in huge maintenance cost savings and reduced failure occurrences.

Uittreksel

Regressie Analise van Caterpillar 793D Trok Enjin Falings Data en Gedokumenterde Toestands Moniterings Data Met Gebruik van die Proporsionele Gevaarkoers Model.

("Regression Analysis of Caterpillar 793D Haul Truck Failure Data and Through-Life Diagnostic Information Using the Proportional Hazards Model")

W.A. Carstens

Departement Bedryfs Ingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MScIng (Bedryfs)

Maart 2012

Fisiese Bate Bestuur (FBB) is besig om 'n groter bekommernis vir maatskappye in die bedryf te word. Die Britse Standaarde Instituut se spesifikasie vir optimale bestuur van fisiese bates en infrastruktuur is PAS55. Volgens PAS55 is FBB die "sistematiese en gekoördineerde aktiwiteite en praktyke wat deur 'n organisasie optimaal sy fisiese bates, hul verwante prestasie, risiko's en uitgawes vir die doel van die bereiking van sy organisatoriese strategiese plan beheer oor hul volle lewensiklus te bestuur".

Een Sleutel Fokus Area (SFA) van FBB is Bate Versorgings Plan (BVP) ontwikkeling. Hierdie is onderhouds strategieë wat bate betroubaarheid verbeter of verseker tydens die volle bruikbare lewe van die bate. Een onderhoud strategie is Toestands Gebasseeerde Onderhoud (TGO) wat besluite baseer op Toestand Monitering (TM) informasie maar neem nie die vorige falingsgeskiedenis van die bate in ag nie. Ander onderhoud strategieë soos Gebruik Gebasseerde Onderhoud (GGO) is gebaseer op historiese falingsdata maar neem nie TM inligting in ag nie.

Regressiemodelle neem beide TM data en historiese falings geskiedenis data in ag ten einde die onderliggende falings gedrag van die gegewe bate te verteenwoordig. Hierdie modelle kan baie nuttig wees vir BVP ontwikkeling te danke aan die feit dat Bruikbare Oorblywende Lewe (BOL) geskat kan word en/of die langtermyn lewenssilus koste geoptimeer kan word.

Die doelwit van hierdie tesis was om historiese falingsdata en TT data goed genoeg te modelleer sodat BOL of optimale langtermyn lewensiklus kostes bepaal kan word om opgeneem te word in BVP ontwikkeling. Hierdie bepalings word dan gebruik in besluitnemings modelle wat gebruik kan word om onderhoud skedules op te stel, d.w.s. om 'n BVP te ontwikkel.

Verskeie regressiemodelle was geëvalueer om die regte model te vind waarmee die doel van hierdie tesis te bereik kan word. Die mees geskikte model vir die navorsingsprojek was die Proporsionele Gevaarkoers Model (PGM). 'n Omvattende ondersoek oor die PGM is onderneem wat fokus op die wiskunde en die praktiese implementering daarvan.

Data is van die Suid-Afrikaanse mynbedryf verkry en is gemodelleer met behulp van die Weibull PGM. Dit was bevind dat die ontwikkelde model resultate geproduseer het wat 'n akkurate verteenwoordinging van realiteit is. Hierdie bevindinge bied 'n opwindende basis vir die ontwikkeling van toekomstige Weibull Proporsionele Gevaarkoers Modelle wat kan lei tot groot onderhoudskoste besparings en minder onverwagte falings.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations:

- Dr. PJ Vlok, my study leader, for his support, dedicated guidance and the time he invested in this research project; which allowed and motivated me to meet ambitious personal deadlines;
- Ernest Stonestreet of Anglo American, for sponsoring me to visit Sishen and obtain data;
- Louis Vlok, for his support;
- Prof. Corne Schutte and Tanya Visser for their supervision and guidance.
- My family, for their continued support;
- Our heavenly Father, for giving me the strength and determination to complete this project.

The Author
December, 2011

Contents

List of Figures

List of Tables

Nomenclature

β	Weibull shape parameter
η	Weibull scale parameter
$\overline{\gamma}$	Row vector containing regression coefficients
$\iota(t)$	Rate of Occurrence Of Failure as a function of time
λ	Functional term dependent on time and covariates
$\overline{z(x)}$	Covariate vector dependent on time
a	Time required to perform preventive maintenance
b	Time required to perform maintenance after unexpected failure occurred
C_{X_c}	Cost of Corrective
C_{X_p}	Cost of preventive maintenance
$C(X_p)$	Total cost per unit time as a function of X_p
E	Expected value
f	Probability density distribution
F	Cumulative failure distribution
h	Force of mortality
H	Cumulative Force Of Mortality
L	Likelihood
m	Total number of observed failures
N(t)	Observed number of failures at time t
n	Observed number of failures
P	Probability
r	Total number of observed events
R	Reliability function
t	Continuous time
T_i	ith Discrete repair time

x	Continuous global time
X_i	ith Discrete event time measured in local time
X_p	Recommended preventive replacement time
X_c	Inspection period
$\tilde{X}_{r+1}$	Upper confidence limit of the Residual Life Estimation
$\underset{\sim}{X}_{r+1}$	Lower confidence limit of the Residual Life Estimation

Glossary

ACP	Asset Care Plans
AFTM	Accelerated Failure Time Model
AHM	Additive Hazards Model
BAO	Bad As Old
BOWN	Bad as Old but Worse than New
CM	Condition Monitoring
CBM	Condition Based Maintenance
CMMS	Computerized Maintenance Management System
DIM	Design Improvement Maintenance
EAMS	Engineering Asset Management System
EHRM	Extended Hazard Regression Model
EHS	Environmental Health and Safety
FOM	Force Of Mortality
GAN	Good As New
HMM	Hidden Markov Model
HPP	Homogeneous Poison Process
IID	Independent Identical Distribution
ISO	International Organization for Standardization
KPA	Key Performance Area
LCC	Life Cycle Cost
ML	Marginal Likelihood
MLE	Maximum Likelihood Estimate
NHPP	Non-homogeneous Poison Process
PAM	Physical Asset Management
PAS55	Publicly Available Specification 55
PF	Potential-Failure
PHM	Proportional Hazards Model
PIM	Proportional Intensity Model
PL	Partial Likelihood
ROCOF	Rate of Occurrence Of Failure

ROI	Return On Investment
RP	Renewal Process
RUL	Residual Useful Life
RULE	Residual Useful Life Expected
TR	Time Ratio
TTT	Total Time on Test
UBM	Usage Based Maintenance
WO	Worse than Old

Chapter 1

Introduction

1.1 Introduction

Industry has only recently started to realize the importance of optimized maintenance activities. According to Mobley (2002), although maintenance is one of the driving factors behind reliable and efficient operation, many industries still knowingly perform ineffective maintenance actions.

Heng (2000) states that the annual maintenance expenditure in American companies in the year 2000 was approximately $1.2 trillion which consisted of roughly 50% wasted expenditure. This was also shown by Heng *et al.* (2009) indicating that up to a third of maintenance expenditure is being wasted on ineffective maintenance actions. Today's complex and advanced machines demand highly sophisticated maintenance strategies in order to sufficiently maintain them.

From this can be seen that a large portion of the total expenditure industry spends on maintenance is being wasted on unnecessary maintenance actions. Therefore a need exists to optimize maintenance actions in order to minimize unnecessary maintenance actions. For this reason there is major scope in the South African industry to apply Physical Asset Management (PAM) principles to improve the overall performance and reliability of assets.

1.2 Problem Statement

An important topic of discussion in industry today is PAM. According to British Standards Institutes' Publicly Availably Specification (PAS55:2008), PAM is the systematic and co-ordinated activities and practices through which an organization optimally manages its assets, and the associated performance risks and expenditures over their life cycle for the purpose of achieving its organizational strategic plan. In most cases managing physical assets such that a desired and sustainable objective is achieved.

In most cases PAM comprises of many different elements such as Work Planning and Control (WPC) and Financial Management (FM). Several other elements are presented in Chapter 2. Most companies develop their own PAM policies to effectively and efficiently manage their own assets. To manage assets effectively and efficiently, it is vital to maintain these assets. Companies typically develop several Key Performance Areas (KPAs) which they apply in order to effectively manage and maintain an asset to fit its' specific functional needs.

One KPA of PAM is Asset Care Plan (ACP) development which consists of several tactical preventive maintenance strategies. Maintenance practitioners mainly consider two preventive maintenance strategies. The first being a strategy which bases its maintenance actions purely on the age of the asset measured in time, kilometers, liters processed, tons or any other process parameter. This type of maintenance strategy is one of the most popular strategies found in industry and is formally known as Usage Based Maintenance (UBM).

In the case of UBM, Condition Monitoring (CM) information is disregarded and maintenance is purely based on the previous failure histories. The determined maintenance instant should not allow that too much residual life is wasted keeping in mind that extending the maintenance instant increases the probability of failure.

The second popular strategy used in industry is when maintenance action is based on CM information in which case maintenance is taken when a certain condition reaches a predefined failure level. This strategy is formally known as Condition Based Maintenance (CBM). This strategy does not take into account the asset's previous failure histories.

Preventive maintenance strategies has the sole purpose of attempting to prevent the occurrence of unexpected failure. Unexpected failures have serious consequences such as damage to the asset, production losses and in some cases loss of life. Compared to the cost of unexpected failures, preventive maintenance is often relatively inexpensive in that many of the unwanted financial consequences are eliminated.

Both of the mentioned maintenance strategies have their own respective advantages. Individually, many studies have been performed to enhance their maintenance decision making capability. However, not much work has been done to incorporate both these maintenance strategies into one maintenance decision making tool. Therefore, the challenge undertaken in this thesis is to investigate models which take into account both CM information and historical failure data which improve practical preventive maintenance decision making. This will be done by using the methodology given in Section 1.4.

1.3 Research Objectives

The objective of this research project is to master the relevant literature in order to gain a detailed understanding of the model ultimately selected to solve the problem described in the problem statement. The researched theory will be validated by applying the model to industry data whereby the accuracy of the estimations made with the model will then be illustrated.

1.4 Research Methodology

In order to achieve the research objectives given in Section 1.3, the research methodology for the following chapters are listed below.

i. Perform a comprehensive literature study on the field of asset management focussing on several maintenance strategies and also failure data analysis. Identify suitable mathematical models with which the research objectives can be achieved.

ii. Evaluate the identified models according to the objectives of this thesis. The most appropriate model will be selected to accomplish the objectives of the thesis based on the evaluation process.

iii. Perform a comprehensive investigation of the chosen regression model focussing on the mathematics and practical implementation thereof. Study decision making models associated with the selected model with which optimized maintenance decisions can be calculated.

iv. Perform a case study whereby the chosen model's theory is applied to data found in the South African industry.

v. Conclude with a summary of the thesis findings and some recommendations for future research.

Each of the listed methodologies are performed in separate chapters whereby a short description is given at the end of each chapter to indicate what was done and whether the stated methodology was followed.

Chapter 2

Literature Review and Contextualization of the Problem

2.1 Introduction

In Chapter 2 a comprehensive literature study is presented focussing on the field of asset management, several maintenance strategies and failure data analysis. The chapter concludes with several identified models which has the potential to accomplish the thesis objectives.

The British Standards Institutes' Publicly Availably Specification's definition of Physical Asset Management (PAM) is repeated here for convenience; PAM is the systematic and co-ordinates activities and practices through which an organization optimally manages its assets, and the associated performance risks and expenditures over their life cycle for the purpose of achieving its organizational strategic plan. This basically means that managing physical assets such that a desired and sustainable objective is achieved. PAS55 was released due to a demand from industry for a PAM standard. In the not too distant future, PAS55 will be issued as an International Organization for Standardisation (ISO) standard.

However, certain considerations have to be taken into account when looking at PAS55. Some PAS55 considerations include: the absorption of time and key resources, management's commitment and involvement and finally coaches and professional support requirements. Benefits associated with PAS55 include: long term organizational support and PAM focus, the holistic and all encompassing broad scope within PAS55, uniform PAM structure, good PAM guidance, marketplace rewards and structure provision for sustainability in the future.

Fogel (2011), a leading South African PAM consultant, states that performance is a metastable condition which requires continuous investment, either in terms of improving or maintaining of assets to do PAM. The task of PAM is important to perform to be competitive in industry today. However, optimal PAM can be quite complicated. This being said, PAM involves balancing different conflicting drivers such as compliance, performance and sustainability. Compliance to environmental laws, quality and safety. Performance specifications such as productivity and sustainability to ensure a sustainable organizational legacy.

As a result, PAS55's focus is to balance these conflicting drivers in order to achieve optimal PAM. Most large companies develop their own PAM policies to manage their assets accordingly to achieve their desired organizational strategic goals. These PAM policies are often developed alongside several Key Performance Areas (KPA) which might include:

- Strategic Management
- Information Management
- Technical Information
- Organization and development
- Contractor management
- Financial Management
- Risk Management
- Environment, Health and Safety
- Asset Care Plan
- Work Planning and Control
- Operator Asset Care
- Material Management
- Support Facility and Tools
- Life Cycle Management
- Project and Shutdown Management
- Performance Management
- Focused Improvement

Strategic management is the incorporation of organizational values, goals and risk policies into the organizational strategic plan. Organizational strategic plans include the asset management policy, strategy, objectives, plans and implementation.

Information management is the integration of documented business data and information with enterprise resource planning and or enterprise asset management systems. Technical information ensures that information is made readily available in a company to ensure effective knowledge management about the operations and processes involving assets. Organization and development focusses on the recruitment and development of resources to correctly and efficiently manage the assets within a company. Contractor management is decision making tool used to indicate whether certain asset management activities have to be contracted out or done in-house. Financial management is a financial information system containing the financial management policies of the assets in a company with a focus on asset valuation and depreciation.

Risk management is risk evaluation of potential risks assets attribute to the overall business objectives. Environment, Health and Safety (EHS) are the integration of these EHS concerns in a company. Asset Care Plans (ACPs) are the activities which maintain the assets of a company to ensure maximum asset utilization and performance, i.e. maintenance strategies. Work Planning and Control is the identification, planning, scheduling, execution and feedback of work associated with assets. Operator Asset Care equips asset operators with the necessary knowledge and skills to operate, prevent, detect and repair asset problems to improve asset performance and utilization.

Material Management is the management of material inventory to optimize the levels such that it reduces the amount of unnecessary stocked items and ensures availability of necessary items to maintain the assets. Support Facility and Tools ensure that the necessary facilities and tools are available to operations to complete tasks to be done associated with the assets. Life Cycle Management is used to analyze the whole life-cycle of the assets from cradle-to-grave taking into account external factors and financial consequences.

Project and Shutdown Management focusses on project initiation, planning, execution and closure. Performance Management is the analysis of strategic, tactical and operational targets attempting to improve each respective asset performance target. Focused Improvement entails daily management of these targets incorporating structured problem solving techniques for each asset.

Each of the 17 KPA's are aligned with PAS55. In most cases it is not possible for an organization to focus on all of the KPA's due to the resources required and the complex nature thereof.

2.1.1 Asset Care Plans

These plans require the implementation of maintenance strategies to improve asset utilization and performance. Moubray (1997) states that maintenance can be divided into four generations. For the *First* generation (up to about 1950) the logic was: "fix it when it breaks". In this generation industry did not have the large amount of mechanized systems as today and downtime had a much less serious effect.

The *Second* generation (1950 to about 1975) was when plant utilization, longer component life and lower component cost started to play a role in company asset performance requirements. This led to complex machine designs and resulted in industry having to increase their dependency on equipment reliability. This brought fourth the introduction of preventative maintenance to reduce unexpected and unplanned failures and downtime.

The *Third* generation (around the mid 1970's), according to Pintelon and Gelders (1992), saw the development of the terotechnology concept and offered a view on maintenance which was a combination of management, financial engineering and other practices applied to physical assets in pursuit of economic life cycle costs. Terotechnology is defined by Kelly and Eastburn (1982) as the management of assets from an economic management point of view. As a result, even more pressure was placed on maintenance due to higher quality and reliability requirements from industry. This meant that the cost of preventive maintenance increased and an alternative maintenance strategy had to be found.

The *Fourth* generation (up to the present day) focuses on the integration of different maintenance activities and production disciplines. It requires cost effective PAM to ensure asset reliability and availability. This can be done by implementation of a suitable ACP.

There are three challenges facing ACP development: cost effectiveness, failure type and maintenance tactic selection. Cost effectiveness has two factors that need to be balanced. The first factor is the cost of preventive asset care attributed to the cost of material and labour. The second factor is penalty costs associated with asset care such as opportunity costs, loss of revenue and recovery costs.

The second ACP challenge deals with two graphs which are failure probability graphs and failure progression graphs. Failure probability graphs indicate the likelihood that a failure occurs at a certain point in time. A typical example of a failure probability curve is the bathtub curve. This curve represents the failure rate of some asset which has three phases during its lifetime. Infant mortality, useful working life and wear-out. Each of these phases are illustrated in Figure 2.1.

During the infant mortality phase, the failure rate of the item decreases. Failure in this phase is attributed to manufacturing and/or design flaws. During its useful working life, the item has a constant failure rate where failures are random and unpredictable. Once the item reaches the wear-out phase, the failure rate increases due to the item reaching the end of its lifetime. Normally a renewal or reconditioning of the item is initiated during this phase to avoid failure.

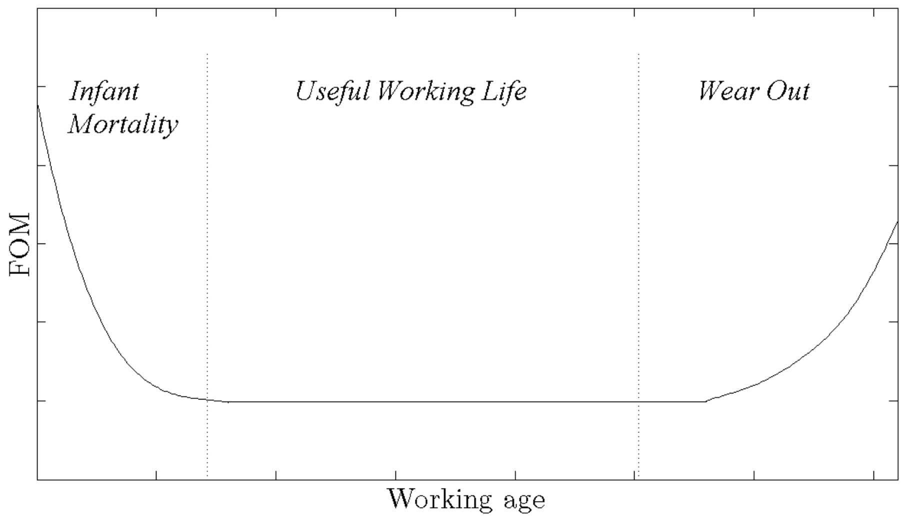

Figure 2.1: Bathtub curve.

A typical example is that of a mechanical system where the failure rate increases with time due to the fact that failure in mechanical systems are most often age related. Another example is that of an electronic system. Electronic systems experience a higher failure rate in the beginning of its lifetime where after it stays constant. Figure 2.1 illustrates these two concepts.

The second type of graphs, failure progression graphs, plot asset condition over working age of the asset. A typical failure progression graph is shown in Figure 2.3. Two points are defined on this graph, Potential failure, P, and Functional failure, F. This is known as the PF interval. Point P indicates where indications of failure start to appear. As time increases after point P, maintenance action has to be taken prior to point F. Point F indicates that a functional failure has occurred. From the graph it can be seen that point P acts as a trigger to initiate some maintenance action before point F to avoid a functional failure.

The third ACP challenge, maintenance tactic selection, involves determining what type of maintenance tactic should be chosen in an attempt to optimally man-

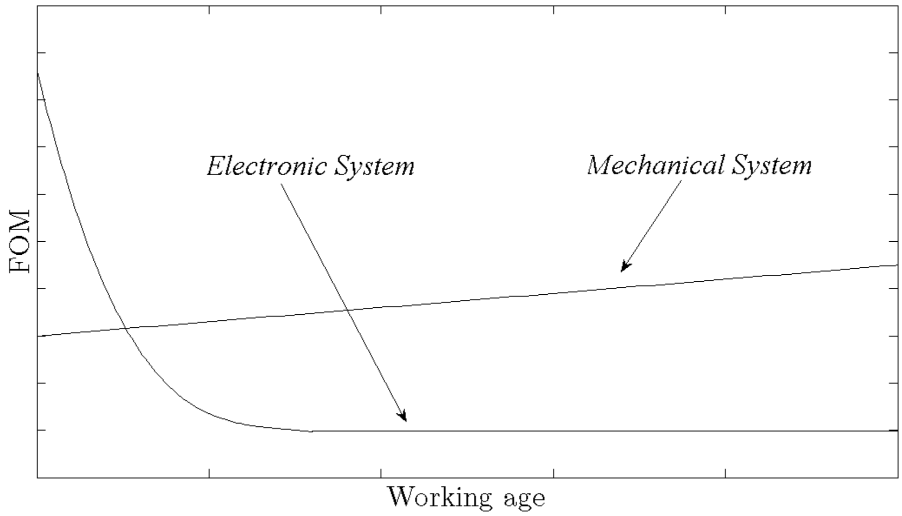

Figure 2.2: Failure rate graphs of typical mechanical and electronic systems.

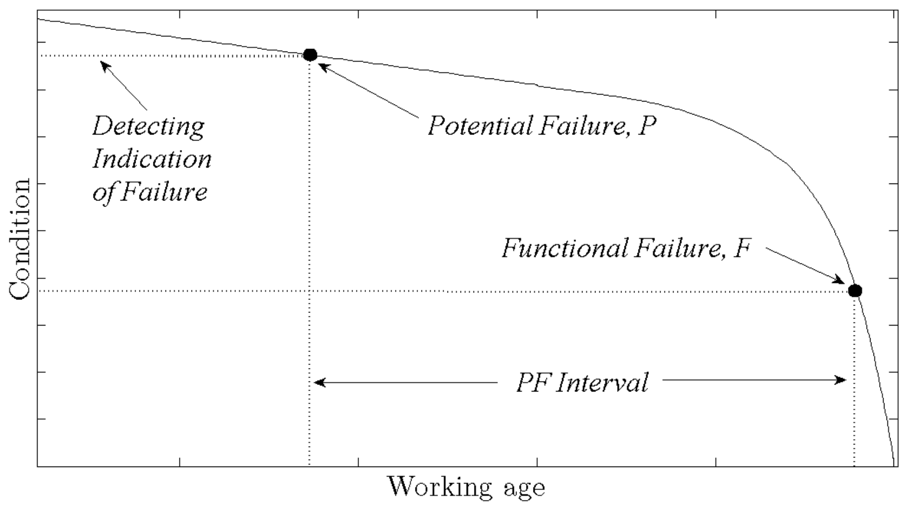

Figure 2.3: Failure progression graph indicating PF interval.

age an asset. Through the years many different maintenance tactics, or strategies, have been developed to keep up with the ever increasing competitiveness of industry.

Two different types of ACPs exist, tactical and non-tactical ACPs. An outlay of the following discussion is given in Figure 2.4.

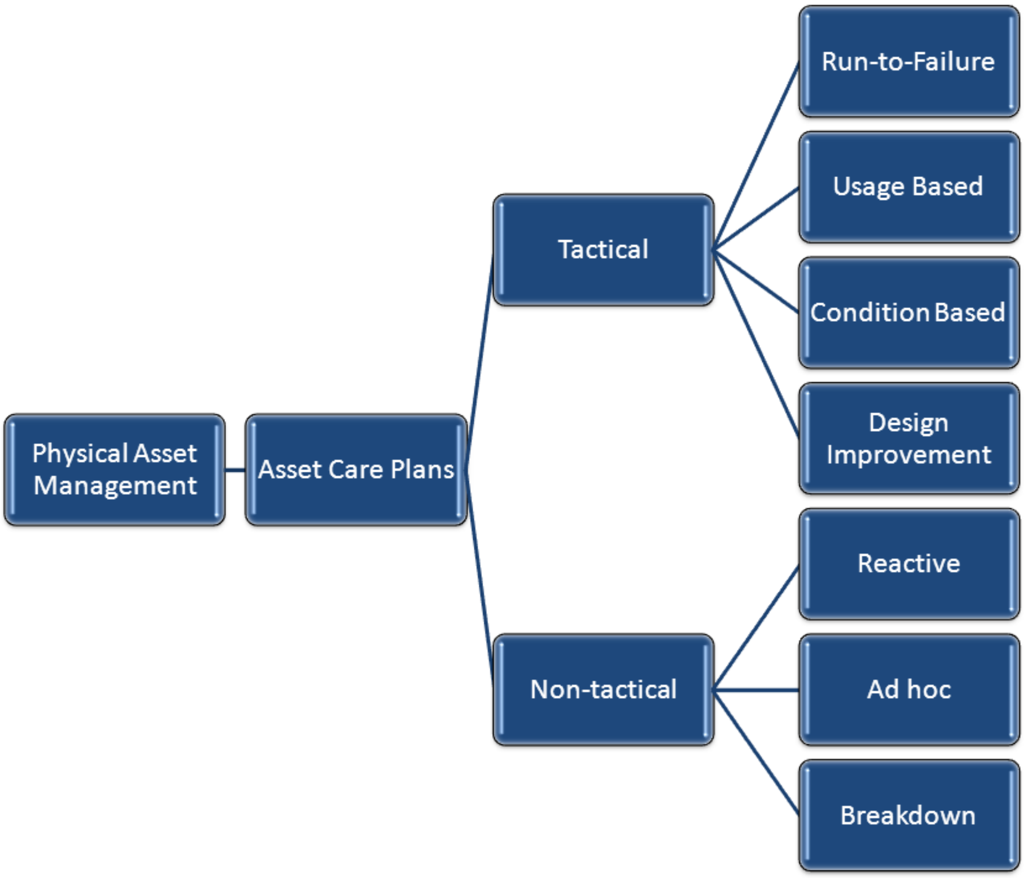

Figure 2.4: Physical Asset Management and the associated maintenance strategies.

Non-tactical ACPs are when maintenance is driven by the assets and assets are operated to failure. Maintenance actions are performed to repair or renew these unplanned failures. Note that in this case failure is unplanned. This means that there are no procedures in place to handle failure once it occurs. Assets are therefore repaired and operated until it fails without taking the necessary precautions to deal with failure.

Unplanned failures result in unreliable processes which cannot be used if optimal asset utilization and performance is wanted. Note that unplanned failures and unexpected failures are not the same. All failures are unexpected. However, unplanned failures can be planned for by putting the necessary contingency plans in place for when a failure occurs. Swanson (2001) indicated that typical examples of non-tactical maintenance strategies are: Reactive, Ad-hoc and Breakdown.

In contrast to non-tactical ACPs, tactical ACPs ensure that maintenance is in charge of the assets. Typical examples of tactical maintenance strategies include: Run-to-Failure, Usage Based, Condition Based and Design Improvement.

Run-to-Failure maintenance Moubray (1997) states that *Run-to-Failure* maintenance operates an asset until it fails and the failed component is then repaired or replaced. This might lead to unplanned downtime which has economic consequences and a tremendous effect on asset reliability and availability. However, *Run-to-Failure* maintenance should have the proper procedures in place to manage the repercussions of an unexpected failure. In this case failure is unexpected, but planned for.

This strategy is linked to the first maintenance generation where manpower and resources are minimized to keep equipment and components running. Disadvantages of this strategy, according to Paz and Leigh (1994), are that it leads to unexpected failures (downtime), reduced asset utilization, poor product quality and high maintenance costs.

Design Improvement Maintenance
Design Improvement Maintenance (DIM) is whereby an asset is modified in attempt to reduce failure and improve asset utilization and performance. This is not a standard maintenance activity action and might be an once off activity or a series of activities to identify an effective new design.

Usage Based Maintenance
Usage Based Maintenance (UBM) is also known as fixed time maintenance or scheduled maintenance. Scheduled maintenance actions are taken at intervals of usage based on the age of the asset and can be done using a process parameter such as days, tonnage handled, kilometers driven, products processed etc. These maintenance actions are executed without considering the condition of the equipment which allows maintenance to be done when it is not necessary. Common maintenance actions and inspections include repairs, renewals and lubrication.

Condition Based Maintenance
This approach is implemented by continuously or discreetly, monitoring equipment to detect worsening conditions to undertake maintenance action in order to restore

equipment to a "good-as-new" state before failure occurs. Maintenance action is taken when the health of the asset deteriorates to a predefined failure level. Performing maintenance in this manner is a preventive strategy which reduces the probability of failure by taking preventive action before failure occurs. Deloux *et al.* (2009) argued that the advantage of this strategy is that maintenance actions are only taken when it is imminent due to the fact that it is based on the physical condition of the asset rather than fixed time intervals or failure.

It is therefore evident that of all the discussed maintenance strategies, CBM attempts to prevent failure whereas the other strategies only focusses on maintaining assets without taking the asset's condition into account. Therefore, CBM has tremendous potential and is discussed in the Section 2.2.

Several authors such as Madu (2000), Nilsson and Bertling (2007), Wang *et al.* (2007), Endrenyi *et al.* (2001) and Márquez *et al.* (2009) indicate how the maintenance strategies presented in this section can be applied in industry.

2.2 Condition Based Maintenance

As mentioned, CBM utilize diagnostic equipment to monitor a condition of an asset and act only when maintenance is imminent before failure occurs. It ensures that maintenance action is taken as a preventive measure. An advantage of CBM is that it is possible to determine whether an asset can operate successfully during a production period whereas an UBM strategy might indicate that a component is due for replacement during the same production period.

Consequently, CBM is the preferred strategy due to the certainty it puts upon the asset's operational capability and the fact that the actual condition of the asset is taken into account. The idea of CBM is to ensure maintenance is done only when it is necessary resulting in a reduction in inventory levels and wasted maintenance actions. Other advantages include an increase in reliability and a reduction in the human error margin due to reduced maintenance operations.

It should however be noted that CBM can only give certainty about asset operation for a short period of time. Another major concern facing CBM is the cost of implementation. Equipment used for CBM are extremely expensive to procure and install and does not guarantee any returns on this investment. Consequently a company needs to evaluate the importance of an asset before implementing CBM equipment.

Information extracted using CBM equipment is known as Condition Monitoring (CM) data. CM data is also known as explanatory variables or concomitant information. The CM data can be categorized into either quantitative or subjective data. Quantitative data include measurements of (but are not limited to): temperature, tribology, vibration, pressure, oil content, stress, etc. Data may be time-dependent or time-independent and can either be measured continuously or measured at fixed process intervals.

Subjective data is known as indicator variables. This might include: whether oil was changed at each service interval, what type of oil used, maintenance team used, type of installation setup and supplier used. Subjective data might be found on job cards which contain information such as who maintained, installed and supplied an item. For example, the inclusion of a subjective data provides the ability to determine whether a maintenance team is associated with an event. It is then possible to track down some miscellaneous cause of a failure.

Subjective data can also be information such as job cards which contain information such as who maintained, installed and supplied an item. The inclusion of a subjective covariate gives the ability to see whether it (absence of an oil change, certain maintenance team, supplier, setup, etc.) is associated with an event. It is then possible to track down a possible cause of an event. With this data it is possible to see whether neglecting maintenance contributes to the occurrence of an event.

It would be ideal to obtain other subjective data by speaking with the technical personnel that maintain and run the items. With their knowledge and experience it might be easier to find subjective covariates which they believe contribute to the occurrence of events.

Unfortunately the recording of CM data is done rigorously (and at great expense) and rarely analyzed to enhance maintenance decision making. With accurate data it is possible to define a failure condition for a certain condition being measured. Once the asset condition reaches this condition, maintenance action should be initiated to repair or renew the asset before failure occurs. In the case where only one condition is being measured, it is reasonably simple to define a failure condition.

Complex and expensive assets measure several different conditions to have as much information about the health of the asset as possible. When several conditions are being measured, failure definition becomes extremely complex. Each of the conditions being measured might lead to failure but analysing one condition is not adequate to prevent failure. Establishing a predefined failure condition when several conditions are being measured is extremely complex due to the intricacy of identifying a correlation between the conditions being measured.

Other challenging factors such as environment and unforeseen- and random-failures increase the complexity of CBM. That being said, Heng *et al.* (2009) stated that effective CBM implementation consists of three key elements.

(i.) Data Acquisition.

(ii.) Data Preparation.

(iii.) Decision Making: Development of asset management policies with information generated by processed asset data.

Each element is briefly discussed in the following sections.

2.2.1 Data Acquisition

Data acquisition in CBM is the collection and recording of useful asset data to be used for maintenance decision making. A requirement of PAS55 is that an organization shall design and maintain systems for managing asset management information. Consequently, employees and relevant persons involved shall have access to any asset information necessary to complete their tasks and activities.

Different methods exist to document this information and forms part of the Information Management KPA. Traditionally job-cards are used to document asset information but software such as Computerized Maintenance Management System (CMMS) and Engineering Asset Management System (EAMS) can be used to simplify this task. Software packages such as this has the capability of linking all an assets' information and storing it at on location. This property allows for decreased data processing duration making data easily available for decision making. Another important capability of these software packages are that it can be used to generate and control work orders.

Asset data includes historical failure data and CM data. Historical failure data contains the recorded time of failure and/or any preventive intervention such as: preventative maintenance, scheduled maintenance and renewal. Any action taken which influenced the asset's failure time is considered to be an event and has to be documented in the historical failure data. The time scale used can be any convenient process parameter such as days, tonnage processed, liters pumped, kilometers driven, products processed, etc.

CM data, or concomitant information, can be categorized into quantitative and subjective data. Concomitant information can also be recorded and stored using CMMS software packages.

Jardine *et al.* (2006) discussed the fact that industry places more emphasis on CM data than event data which leads to neglected and inaccurate event data. Event data and CM data both play crucial roles in effective CBM. A possible reason why event data is being neglected and of bad quality, is that as suppose to CM data, which is recorded using diagnostic equipment, event data is entered manually. Manual entry of data has a high risk due to the human error involved.

2.2.2 Data Preparation

Data preparation is a process of extracting and selecting the most useful information contained in raw asset data. Data preparation basically entails data cleaning. Data cleaning is the removal of errors from a data set. Countless errors can be found in recorded event data and CM data. Errors in event data might be due to the high margin of human error involved. Typical examples of errors in CM data include faulty equipment recordings, missing data and incorrect data labels. Data cleaning plays an important role in data preparation and further discussion is done in Chapter 4.

2.2.3 Decision Making

Certain forms of CBM analyze event data in conjunction with CM data. Heng *et al.* (2009) states that these forms of CBM use event data and CM data to develop mathematical models to represent the failure (or fault) behaviour of an asset. Mathematical models representing the failure behaviour of assets can the be used to predict the future condition of the asset. Decision making in CBM can either be done with diagnostics or prognostics. These concepts are designed to improve decision making related to maintenance in attempt to maximize asset utilization and performance. Both these concepts require some form of data analysis and model building.

2.2.3.1 Diagnostics

According to Caesarendra *et al.* (2010), diagnostics is the measurement of an asset's condition to identify, detect and isolate a fault condition before failure occurs. Venkatasubramanian *et al.* (2003*b*) presented three diagnostic approaches which are quantitative model-based, qualitative model-based and process history based.

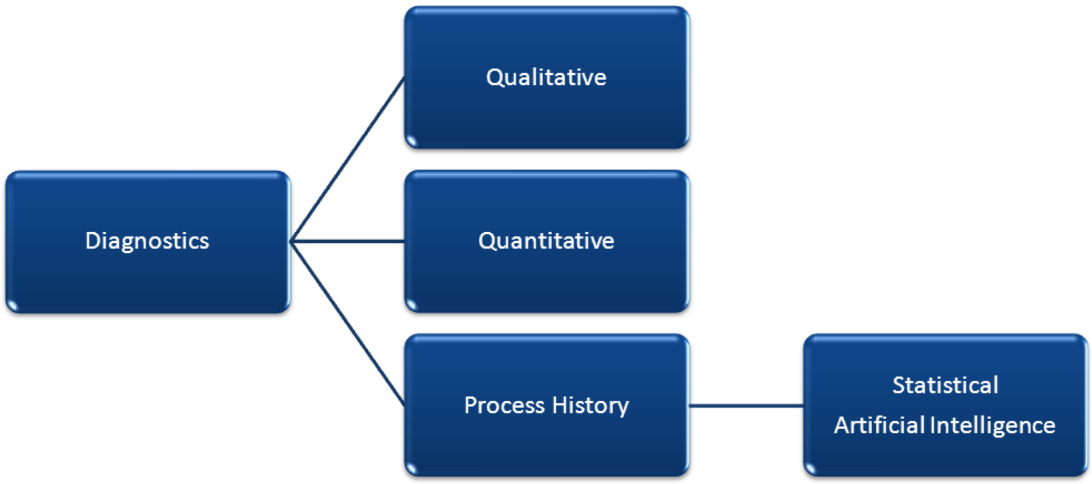

Figure 2.5: Diagnostics and its elements.

Quantitative and Qualitative Model Based: Venkatasubramanian *et al.* (2003*b*) showed that the most popular quantitative model based methods are diagnostic observers, parity relations and Kalman filters. Qualitative model based applications can be seen in Venkatasubramanian *et al.* (2003*a*). Quantitative and model based applications are given by Isermann (2005) and Zheng *et al.* (2006), respectively.

Process History Based: The first process history method is statistical based. A statistical analysis is used to focus on the development of a fault hypothesis by comparing the current system condition with the fault hypothesis, which is a predefined failure condition after which deciding whether a fault is present or not. Ma and Li (1995) showed that bearings experience localized defects and these defects comprise two alternating zero mean Gaussian distributions with their own respective variances. Each Gaussian distribution is known as a hypothesis. By using both variances and a likelihood ratio test, the localized defect is present if the likelihood is greater than a certain defined level.

Kim *et al.* (2001) did a study on an internal combustion engine's air and fuel system. Two hypothesis were developed and thresholds were determined for each hypothesis respectively and then compared to measured values. The thresholds were used to minimize the probability of incorrect fault detection and as an alarm. If the threshold criteria was met, a fault was present.

The second process history method is artificial intelligence based. Sorsa *et al.* (1991) showed that neural networks could be used to mimic human knowledge and pattern recognition ability. Another popular technique discussed by Mechefske (1998) was Fuzzy logic. The study was done using bearing frequency data and different fuzzy logic membership curves were applied to find best fault classification method.

2.2.3.2 Prognostics

Prognostics is used to predict when failures and/or faults may occur and to estimate the remaining time left before these failures and/or faults occur. The advantage of knowing when faults and/or failures occur equips maintenance teams with the knowledge to reduce maintenance costs, improve asset utilization and performance by developing accurate maintenance schedules.

Prognostics focusses on two main outcomes. The first outcome is to determine how much time is left before failure and/or fault(s) occur. This is referred to as the Residual Useful Life (RUL) of an asset. The second outcome is the expected time until the next failure and/or fault(s) occur. This concept can be referred to as the Residual Useful Life Expected (RULE) of an asset.

Both RUL and RULE base their estimations on the given CM data and historical failure data. Statistical regression models are suitable for the combined analysis of historical failure data and CM data to determine these estimations. These time-dependent regression models can find a relation between the failure probability using both the historical failure data and CM data. These models enable the prediction of future asset condition at a specified time in the future enabling the estimation of RUL and RULE.

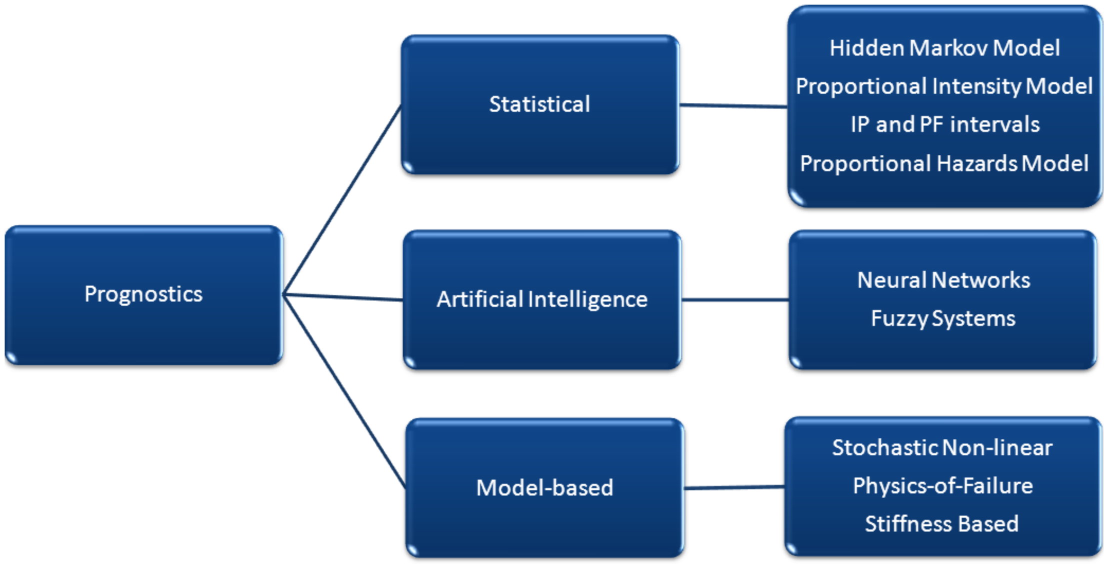

Figure 2.6: Prognostics and its elements.

In prognostics, it is necessary to have the necessary knowledge to represent the failure behaviour or mechanism available. Failure mechanism has two definitions:

(i.) Failure solely depends on the CM data which reflect the actual fault level and predefined condition boundary. Therefore failure occurs if the predefined condition boundary level is reached.

(ii.) A model is developed for the failure mechanism using both the event data and CM data. Different failure definitions can be used such as unsatisfactory performance level, renewal and repair.

Different approaches to prognosis exist and include *statistical*, *AI* and *model-based* approaches.

Statistical Approach: Wang *et al.* (2000) modelled CM data and event data using Gamma distributions to estimate the RUL. A decision model was developed to compare corrective-maintenance and preventative-maintenance duration to decide when maintenance should be done.

Kwan *et al.* (2003) used the Hidden Markov Model (HMM) and trained a HMM model to detect certain faults before failure. HMM is where the system observed is assumed to be a Markov process with hidden states. It is used to model outputs that are generated by a finite number of states. The states are hidden (not visible), the the outputs can be measured. With the knowledge of the outputs it is possible to determine the hidden states of the system.

Chinnam and Baruah (2003) showed how HMM clustering methods can be used to facilitate autonomous prognostics by capturing the transition health states of each stage of the failure process by fitting a Gaussian distribution and estimating the RUL. Wang (2007) indicated how CM data can be used to identify the two stages of a components life, normal operating stage and failure delay stage. This was done by modelling the unobserved state transitions using a HMM and comparing results by fitting Weibull and Gamma distributions.

Goode *et al.* (2000) indicated how machine life can be separated into two intervals: IP (Installation Potential) interval in which the machine is running without faults and the PF (Potential Failure) in which the machine is operating with a fault. Weibull distributions were assumed for the IP and PF interval respectively and failure prediction was derived for each interval and the RUL estimation was done.

Vlok *et al.* (2004) showed how the Proportional Intensity Model can be used to estimate bearing RUL. You *et al.* (2010) showed how the Proportional Hazards Model (PHM) and a two-zone extension thereof can be used to estimate the RUL. Banjevic and Jardine (2006) discussed the estimation of RUL using the PHM in conjunction with a Markov failure time process. Both PIM and PHM are regression models are frequently used for reliability applications.

AI Approaches: These approaches found in literature were neural networks and fuzzy systems. Neural networks are developed by algorithmic training and fuzzy systems utilize linguistics for future system condition prediction. Wang and Vachtsevanos (2001) proposed a prognosis architecture using a wavelet neural network to estimate fault growth and RUL. Wang *et al.* (2007) used a fuzzy analytic hierarchy process to compare and evaluate different maintenance strategies.

Wang *et al.* (2004) showed how fuzzy systems can be trained using neural networks and to determine failure definitions thereof. Chinnam and Baruah (2004) showed how a neuro-fuzzy model can be used to determine reliability estimation for individual components using expert linguistic opinion and empirical data.

Model-based Approach: Model-based approaches require dynamic and/or static information of an asset to develop a mathematical model in order to estimate RUL.

Ray and Tangirala (1996) applied a non-linear stochastic fatigue crack dynamic model to on-line monitoring of accumulated degradation and current degradation rates. Li *et al.* (2000) developed a stochastic non-linear bearing fatigue defect-propagation model which was recursively applied without prior knowledge of the system. Kacprzynski *et al.* (2004) showed how physics-of-failure modelling fused with diagnostic information and could be utilized to enhance the accuracy of helicopter gear RUL estimation.

The study by Qiu *et al.* (2002) was done to develop a stiffness-based prognostics model based on dynamic on-line vibration analysis and accumulative degradation mechanics. The study indicated that bearing lifetime can be analyzed and predicted by monitoring the change in dynamic system stiffness based on the on-line vibration measurements. Caesarendra *et al.* (2010) used a model in the developing stages utilizing Monte Carlo based techniques and is known as the Particle Filer method. It was shown that it exhibits some potential in predicting trending data in non-linear systems.

2.2.3.3 Conclusion

Diagnostics and prognostics are both CBM approaches used to attempt to prevent failure. Diagnostics take into account CM data and tests whether a predefined failure condition is reached. If this condition is met, maintenance action is initiated. Prognostics combines both CM data and historical failure data to estimate asset future condition. A major drawback of diagnostics is that it can only take into account one condition to define the failure condition.

Consequently, diagnostics is limited by the fact that it cannot take into account several asset conditions to detect failure. This leads to less accurate results compared to prognostics. Three prognostic approaches were discussed in Section 2.2.3.2. Statistical approaches fit distributions to data and develop models to predict RUL and RULE. AI approaches train algorithms how assets behave which are then used to predict asset RUL and RULE. Model-based approaches model data using existing physics models.

From what was found in literature, statistical approaches have widely been applied in the field of reliability in many different industrial applications. Statistical approaches are extremely flexible and has many applications.

2.3 Failure data analysis

Failure data analysis is the statistical analysis of item/system failure data and the point of interest is the time until failure occurs. Time is a process parameter defined on any convenient time-scale. This process parameter is defined from a certain known origin up to the occurrence of an event event. Therefore it is important to define how event times are measured. Event arrival times are typically measured as shown in Figure 2.7.

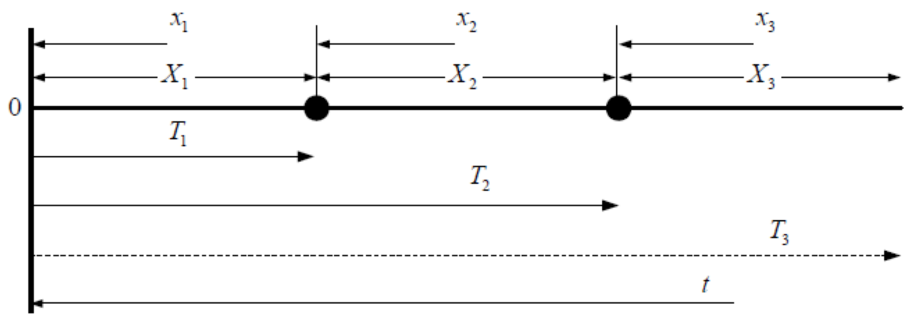

Figure 2.7: Failure time measurement (Adapted from Vlok (2001)).

In figure 2.7 X_i refers to the interarrival time between the $(i-1)^{th}$ event and the i^{th} event where $i = 1, 2, .., n$. The interarrival event times are random variables with $X_0 \equiv 0$. This variable is also known as *local* time. The so called real variable, x_i, is used to indicate the elapsed time following the most recent event.

T_i refers to time measured from 0 to the i^{th} event time. This variable is referred to as the *arrival* time to the i^{th} event.In most cases X_i are used to analyze non-repairable systems whereas T_i is used to analyze repairable systems. The overall time scale t is known as *global* time.

As shown by Coetzee (1997), after the event time has been measured it is necessary to determine the system type. Two system types exist which are repairable and non-repairable. A *repairable* systems can be repaired or renewed to its full potential to perform its intended functions correctly without the replacement of the entire system.

Time between events in repairable systems are dependent and not from the same distribution, which means that repairable systems failure data is dependent and not identically distributed. With *non-repairable* systems, maintenance actions are used to renew an entire system when a failure occurs.

Renewal results in an entire system restoration to a new state. Failure data obtained from non-repairable systems are independent and identically distributed (i.i.d). After a system has failed and action is taken to repair it, it is in one of the following states:

(i.) As good as new (GAN): A perfect repair or renewal which restored a system to a condition equivalent to a new system.

(ii.) As bad as old (BAO): When a minimal repair was done which restored the system to the same state just before the event.

(iii.) Better than old but worse than new (BOWN): A repair restoring the system to a state better than just before the event but worst than a total renewal, basically a state between old and new.

(iv.) Worse than old (WO): A repair or renewal which restored a system to a condition worse than before the event.

Further discussion on these assumptions are done in Section 2.3.3. An important point of interest is model type selection. Selecting the most appropriate model to represent the failure behaviour of asset data is often neglected. Therefore a thorough discussion on this topic is done in the following section.

2.3.1 Model type selection

To model failure data, it is necessary to follow a selection process to determine which model fits the selected data best. It was found that most failure data analysis applications do not pay much attention the model type selection process. Ascher and Feingold (1984*b*) set out a model selecting procedure using various trend tests and dependency tests. Coetzee (1997) and Louit *et al.* (2009) modified this procedure mainly focussing on homogeneous- and non-homogeneous poison process models.

Their procedure was also modified and the resulting failure data model type selection framework can be seen in figure 2.8. The main focus of this framework is on the Non-homogeneous Poison Process (NHPP) and the Renewal Process (RP). With that said, the framework helps to identify the failure process of the system being looked at. The two failure systems include a *non-repairable* system, RP, and *repairable* systems.

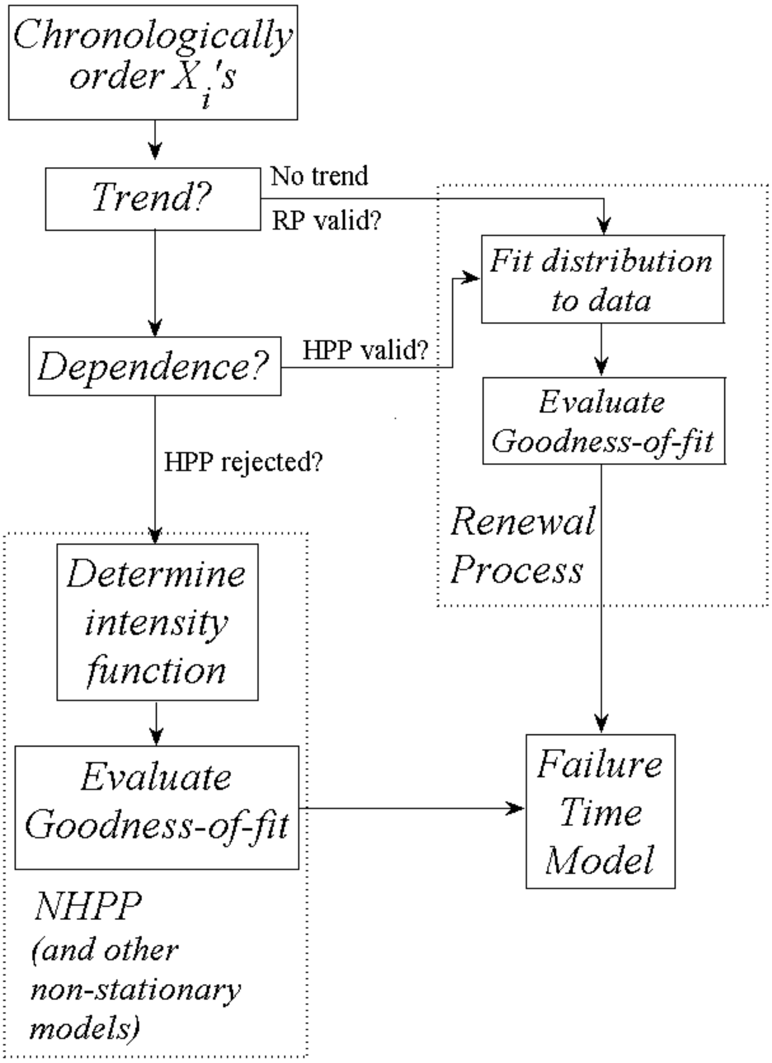

Figure 2.8: Framework for statistical failure data analysis. (Adapted from Louit *et al.* (2009) and Coetzee (1997)).

(i.) *Chronologically order $X_i^{\prime}s$*: The first step in this process is to order the events in their chronological order. Chronologically ordering the events is done to recognize trends in the data.

(ii.) *Trend tests*: Many techniques exist for trend testing. These tests basically tests the renewal assumption. If a trend is present in the data, the renewal assumption is validated and the events are independent and identically distributed. In testing for a trend, the objective is to determine whether a NHPP or Homogeneous Poison Process (HPP) model is applicable to use. Kvaloy and Lindqvist (1998) states that when testing for a trend special care should be taken when selecting a time scale. The selection of the time scale may have an effect on the occurrence of the event pattern. A trend in failure data can be monotonic or non-monotonic.

Louit *et al.* (2009) states that monotonic trends can either be a "happy-system" or a "sad-system". A "happy-system" trend exists when the time between the events increase with increase in time. Thus as time increases, the trend decreases. A "Sad-system" trend exists when an increase in failure occurs as time increases. In this case the trend increases as time increases.

Non-monotonic trends follow the bathtub curve and are therefore time dependent. Thus time between failures decrease in the beginning, then stays constant and decreases near the end. A common graphical technique used for trend testing is the Total Time on Test (TTT) plot.

Analytical trend testing techniques include: Laplace test, Lewis-Robinson test, Anderson-Darling trend test and Military handbook test. The Laplace trend test was developed by Laplace (1773) and application of these graphical and analytical methods can be seen in Kvaloy and Lindqvist (1998). However, only the Laplace and Military handbook tests have a HPP null hypothesis. This means that if no trend is found in the data, an HPP can be applied because the data is i.i.d. If however a trend is found in the data, the data is assumed not to be i.i.d. and a NHPP can be used.

(iii.) *Dependence Tests*: The test for dependence is done to test for dependency between failure times without a long term trend. Coetzee (1997) suggests that any serial dependency test is adequate and proposes a test presented by Krishnaiah and Sen (1984).

(iv.) *Renewal Process*: If the process leads to the RP, the failure data is i.i.d. and therefore generated by a renewal process. Modelling this data is done by fitting a suitable statistical distribution to the data.

2.3.2 Censored data

Data is censored when an asset being monitored is taken out of service or is still operating correctly when the end of the measured interval is reached. According to Martinussen and Scheike (2006), right censoring is when an asset operated acceptably for longer than the observed right censoring value or otherwise, the measured interval.

Right-censored failure data can however be used to estimate the regression parameters that relate to the failure behaviour of the asset under study. However, if the censored data is not taken into account, basing the statistical analysis on only the complete data would give bias results. Additionally, censored data cannot be used in trend testing as it does not represent true failure times and could lead to erroneous results.

2.3.3 Reliability Theory

Reliability theory is focussed on estimating the probability or reliability of a certain item based on historical failure data. Reliability is defined as the probability that an item correctly performs its functions when installed in its intended application at time x.

$$\text{Reliability}=\text{P(No failure at time } x)$$

Note, Nachlas (2005) showed that it is possible to model stochastic survival times of identical devices with probability and hence by a cumulative density distribution. This distribution is the basis of four important descriptions of reliability.

(i.) *Probability density function*, *f*, which is the probability that an item that has not failed in the interval $(0, x)$ fails in the interval $(x, x + \delta x)$.

(ii.) *Cumulative failure distribution* or *Survivor function*, F, which is the probability that an item fails in the interval $(0, x)$.

(iii.) *Reliability function*, R, which is the probability that an item does not fail in the interval $(0, x)$.

(iv.) *Force of mortality* is the *instantaneous failure rate* of an item that is working correctly at time x in the interval $(x, x + \delta x)$.

These functions form the base of reliability engineering and are all related in some way in that if one function is known, it is possible to derive the other three. To illustrate this, the four functions is briefly discussed below.

2.3.3.1 Probability Density Function

The probability density function, $f(x)$, represents the probability that an item that has not failed in the interval $(0, x)$ fails in the interval $(x, x + \delta x)$ at time X. It is given that:

$$f(x)dx = P(x \leq X < x + \delta x) \tag{2.3.1}$$

with $f(x) \geq 0$ and the area under the probability density curve equal to unity, $\int\limits_0^\infty f(x)dx = 1$.

2.3.3.2 Cumulative Failure Distribution

The cumulative density distribution, $F(x)$, is the probability of a failure occurring before x, $F(x) = P(X \leq x)$. This probability can be obtained using the accompanying probability density function and is given by:

$$F(x) = \int\limits_0^x f(x)dx \tag{2.3.2}$$

As x tends to infinity, $F(x)$ tends to unity.

2.3.3.3 Reliability Function

The reliability function is complimentary to the cumulative failure distribution and is also known as the survivor function. The reliability function, $R(x)$, is the probability that an item survives up to x, $R(x) = P(X \geq x)$. The reliability function is given by:

$$R(x) = \int_x^\infty f(x)dx \tag{2.3.3}$$

where $R(x) = 1 - F(x)$. As x tends to infinity, $R(x)$ tends to zero.

2.3.3.4 Force of Mortality

The force of mortality (FOM), $h(x)$, is the statistical characteristic often used in reliability studies. The probability that an item might fail in the interval $(x, x+\delta x)$, given that it is working correctly at time x, is known as a conditional probability. To define conditional probability, suppose there are two events, A and B, where event A is dependent on event B. The conditional probability of event A given event B is given by:

$$P(A|\, B) = \frac{P(A \cap B)}{P(B)} \tag{2.3.4}$$

where $A \cap B$ indicates that both event A and B take place and is valid if $P(B) \neq 0$. This implies that the conditional probability of event A given event B is the probability that event A and B occurs divided by the probability that event B occurs. Suppose $h(x)$ represents the FOM and $h(x)dx$ the probability that the item is in a failed state at time $X < x + dx$ given that no failure occurred at $X = x$. Therefore, by the definition of conditional provability, the FOM can be given by:

$$h(x)dx = P(X < x + \delta x|\, X > x) = \frac{P(X < x + \delta x) \cap P(X > x)}{P(X > x)} \tag{2.3.5}$$

The numerator indicates the probability that an item might fail in the interval $(x, x+\delta x)$ given that no failure occurred at $X = x$ which is the probability density function, $f(t)$. The denominator indicates the probability that an failure might not occur in the interval $(0, x)$ which is the reliability function, $R(x)$. It can then be shown that Equation 4.6.5 can be given by:

$$h(x) = \frac{f(x)}{R(x)} \tag{2.3.6}$$

Additionally, it is possible to determine the reliability function with the following relation:

$$R(x) = \exp\left(-\int_o^x h(x)dx\right) \tag{2.3.7}$$

The FOM is frequently used to describe the failure behaviour of an item. In many practical applications the FOM of similar complex items follow the bathtub

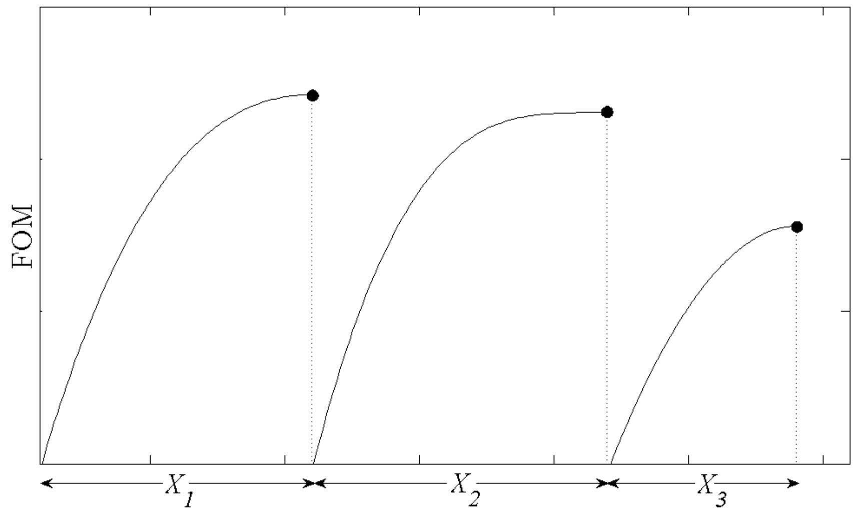

Figure 2.9: Graphical illustration of GAN assumption.

curve. As shown in Figure 2.1, the bathtub curve has three phases. Infant mortality, useful working life and wear-out.

A major concern to maintenance planners is which maintenance strategy to adopt in each of these three phases. When Equation 2.3.6 produces a constant risk, failure is a random occurrence due to a constant failure probability. For constant failure probability items, the run-to-failure maintenance strategy can be used. For a decreasing FOM, the run-to-failure maintenance strategy should also be used due to the fact that the failure probability is decreasing.

However, if Equation 2.3.6 produces an increasing risk, the probability of failure increases with an increase in time. Preventative maintenance has to be considered. An increasing FOM corresponds to the wear-out phase and a preventative maintenance strategy such as CBM definitely has to be considered.

After a maintenance strategy has been selected, it is necessary to define the state of the system after maintenance has taken place. From Section 2.3.1 it was possible to determine whether the observed system is non-repairable or repairable. Knowing the system type presents the state of the system after a renewal or repair. Non-repairable systems are in a GAN state after each renewal. This assumption implies that the FOM is zeroed after each renewal. The GAN assumption is illustrated in Figure 2.9.

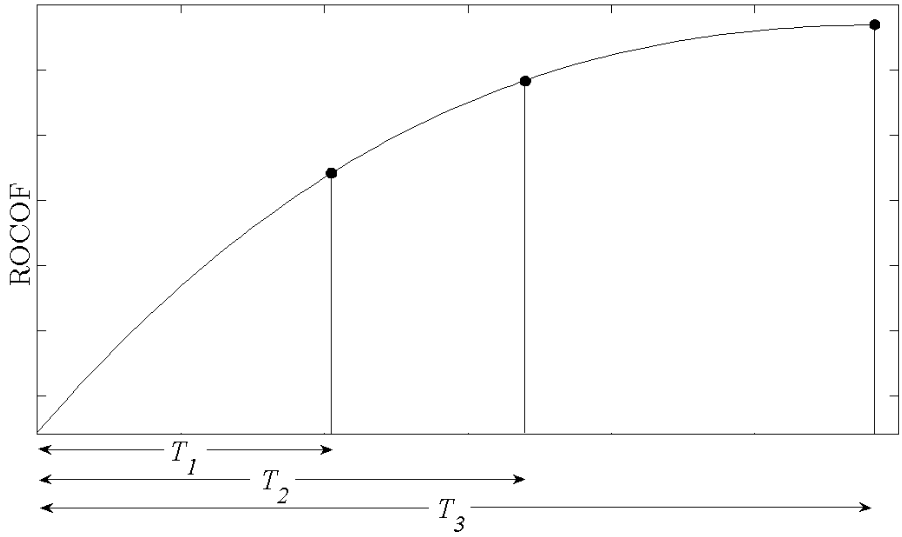

Figure 2.10: Graphical illustration of BAO assumption.

Repairable systems is in a BAO state after a repair. With repairable system modelling it is assumed that simultaneous failures cannot occur. It should be noted that the FOM is not used with repairable systems, but rather Rate of OCcurrence Of Failure (ROCOF). ROCOF is defined as:

$$ROCOF = \frac{d}{dt}E\left\{N(t)\right\} \tag{2.3.8}$$

where $N(t)$ is the observed number of failures in $(0, t]$. Repairable systems can undergo degradation combined with an increased ROCOF causing a worsening trend or even complete renewal of a system. Figure 2.10 is given to illustrate the BAO assumption.

Next we refer back to the wear-out phase of the bathtub curve. When a non-repairable system reaches the wear-out phase, a renewal brings the system back to a GAN state which means that the failure probability is zeroed. Consequently, non-repairable systems can be operated due the state after renewal. However, repairable systems which are repaired in the wear-out phase might experience an increase in failure probability. Renewal therefore has no effect on the system failure probability and system should be renewed to decrease the occurrence of failure.

Estimating the four reliability functions can be done using the statistical approaches discussed in Section 2.2.3.2. The reliability functions is used to predict asset RUL and RULE. It is possible to use these statistical approaches to model repairable and non-repairable systems.

The only approach looked at in this thesis is regression modelling. This approach is frequently used in failure data analysis due to the fact that it is possible to include time continuous explanatory variables, also known as regressors, or CM data. Regression modelling is a statistical method and is part of prognostics.

Improved asset management can be done by implementing one or more KPA such as ACP development. These ACPs attempt to improve asset utilization and performance through optimized maintenance activities such. Optimizing maintenance can be done by using different types of maintenance strategies. It was shown that some maintenance strategies, such as CBM, poses asset future condition modelling. Prediction of asset future condition is crucial to improve asset management. Most traditional reliability modelling procedures only incorporate historical failure data. On the other hand, most maintenance strategies only take into account CM data.

CM data change stochastically and may influence and/or indicate the survival time of an asset. Omitting CM data might lead models which does not have the ability accurately predict asset future condition resulting in the occurrence of failure. Therefore, a need exists to model both event data and CM data together. Balakrishnan and Rao (2004) stated that models capable of this are known as regression models.

2.3.4 Regression Modelling

Many different regression models can be found in literature. It should however be noted that a number of fundamental regression models exist upon which many variances thereof are based. These fundamental models are modified to suite the needs of particular applications. Selection of regression model to research in this thesis are evaluated according to the following criterion:

1. Previous application in the reliability engineering
2. Model originality

Six regression models were identified namely; Proportion Odds Model, Additive Hazards Model, Proportional Hazards Model, Accelerated Failure Time Model, Proportional Intensity model and Extended Hazard Regression Model. Further discussion of these models are done in Chapter 3.

2.4 Conclusion

This chapter attempted to summarize the first item of the research methodology. PAM and ACP development was introduced where after several maintenance strategies were looked at. It was found that CBM is the most suitable maintenance strategy to achieve the objectives of this study. Further research of CBM indicated that a sub-field thereof is prognostics which is used for RUL predictions.

In order to determine these estimations, a school of thought called failure data analysis was presented. This lead to the identification of six suitable regression models for this study. In Chapter 2 these regression models are evaluated according to the objectives of this research study.

Chapter 3

Regression models

3.1 Introduction

The purpose of Chapter 3 is to evaluate the regression models identified in Chapter 2 according to the thesis objectives. The most appropriate model is then selected to accomplish the desired objectives of this thesis.

The six regression models identified in Chapter 2 were: Proportional Odds, Additive Hazards, Proportional Hazards, Accelerated Failure Time, Proportional Intensity and Extended Hazard Regression. These models were identified based on the criterion given in Chapter 2 which is repeated here for convenience.

1. Application in the field of reliability
2. Model originality

Several models were evaluated according to these criteria whereby six models were identified with the potential to achieve the thesis objectives. Many variations of the six model can be found in literature each being modified for its specific application. A decision was then made to focus on the original six models upon which the variances are based upon.

The methodology of Chapter 3 is to evaluate the identified regression models according to the objectives of this thesis. After each model has been evaluated they are compared whereby the most suitable model is chosen.

Due to the mathematical content in this chapter, a number of mathematical concepts have to be defined. Note that it is assumed that an “event" and a “failure" have the same meaning and can interchangeably be used. Now, consider failure

Table 3.1: Regression model evaluation matrix.

Criteria	**Weight**
Future potential	2^1
Implementation intricacy	2^2
Reliability application	2^3
Flexibility	2^4

time $X > 0$ and a vector of m covariates $\overline{z(x)} = [z_1(x), z_2(x), z_3(x), ..., z_m(x)]$ has been observed. Each $z(x)$ represents an asset condition as a function of time. For generality, assume that covariates are time-dependent even though covariates can be time-independent. Consequently, it is also assumed that the process parameter used in this chapter is *time*.

Furthermore, to incorporate the influence of covariates, a regression coefficient is associated with each of the m covariate vectors. Consider a row vector, $\overline{\gamma}$, of m regression coefficients i.e., $\overline{\gamma} = [\gamma_1, \gamma_2, \gamma_3, ..., \gamma_m]$. Estimation of these regression coefficients are done through model fitting procedures described in Chapter 4.

In this chapter the most suitable model for the thesis is determined. Each model is presented mathematically and graphically followed by its advantages and disadvantages. At the end of each section, the model is evaluated using the evaluation criterion given in Table 3.1 followed by a brief description of each criteria.

Future potential evaluates possible future-use of the model in the field of reliability. *Reliability application* identifies previous applications of the model in reliability. *Implementation intricacy* looks at the complexity of implementing the model and the associated procedures involved. *Flexibility* is used to determine the extent to which the thesis objectives can be met with the particular model.

Each model is given a mark out of five for the respective criteria. To scale the importance of the criterion, a weight is given to each criteria as shown in the second column of Table 3.1. In the final section of this chapter, a summarizing table is given containing all the marks given to each model. This table also contains the total mark of each model thereby indicating the most suitable model for this thesis.

3.2 Proportional Odds Models

The Proportional Odds Model (POM) was introduced by McCullagh (1980) to analyze failure times in the field of biomedicine. Bennett (1983*a*) first described the model in a semi-parametric framework and was further developed by several authors such as Yang and Prentice (1999).

In the POM, covariates act multiplicatively on the *odds* against failure time. Therefore, the *odds* of failure changes when covariate values change. However, as time increases the influence of the covariates diminish. Consider three FOM functions $h_1(x, z_1)$, $h_2(x, z_2)$ and $h_3(x, z_3)$, where each FOM function is influenced by a different covariate. Figure 3.1 indicates how these functions converge illustrating how the influence of covariates diminish over time.

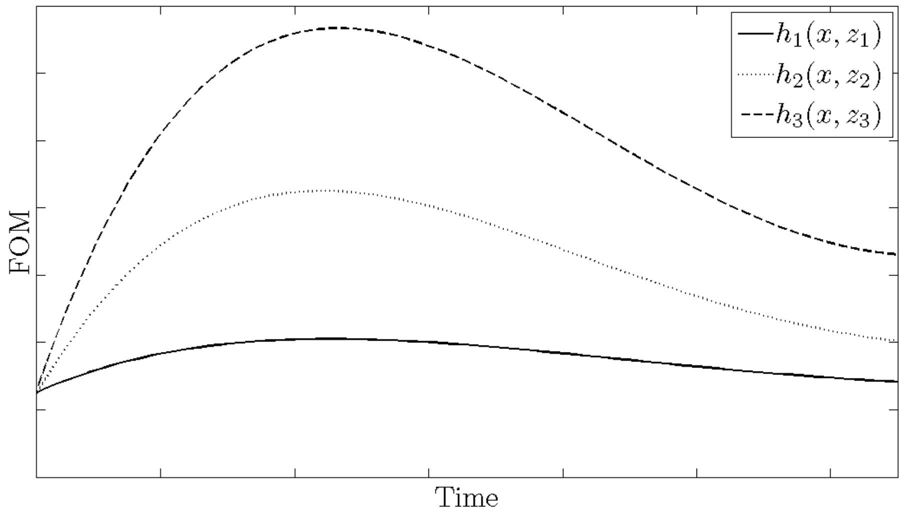

Figure 3.1: Graphical illustration of how the covariate influence diminish over time.

A typical example of systems where the influence of covariates diminish over time are electronic systems. In Chapter 2 it was shown that electronic system often experience decreasing probability of failure during infant mortality and converge as time increases. Consequently, failure is a random occurrence and cannot be correlated to the influence of the covariates. Due to this property, the POM has not seen much application in the field of reliability as reactive based maintenance is applied where systems are experiencing random failure patterns. The *odds* of an individual failing at time x is defined as:

$$\frac{F_X(x)}{R_X(x)} = \frac{1 - R_X(x)}{R_X(x)} \tag{3.2.1}$$

However, by including covariates in Equation 3.2.1, the POM can be given by:

$$\frac{1 - R(x, \overline{z(x)})}{R(x, \overline{z(x)}} = \psi \cdot \frac{1 - R_X(x)}{R_X(x)} \tag{3.2.2}$$

where ψ is a factor representing the multiplicative effect covariates have on the odds of failure. For instance, if ψ increases, the probability of failure is greater which results in a shorter failure time. It should however be noted, covariate inclusion does not necessarily result in a greater ψ value. Covariates can increase or decrease the ψ value.

A commmon form of ψ given by Lu and Zhang (2007) is $\psi = \exp(\overline{z(x)} \cdot \overline{\gamma})$. Furthermore, Yang and Prentice (1999) states that if X is absolutely continuous, the FOM can be given by:

$$h(x, z) = \frac{1}{\psi + R_X(x)} \cdot \frac{dR_X(x)}{dx} \tag{3.2.3}$$

where $R_X(x) = F(x)/(1 - F(x))$. Note that for any two covariates $z_1(x)$ and $z_2(x)$, $h_1(x; z)/h_2(x; z)$ tend to unity with increase in x due to the covariate diminishing property.

Bennett (1983*b*) proposed a profile likelihood method for model parameter estimation. Profile likelihood, also known as full likelihood, estimates the unknown baseline FOM function and the regression coefficients. Murphy *et al.* (1997) applied the profile likelihood method and presented its POM parameter estimation capabilities. Parameter estimation can also be done using Maximum Likelihood Estimation (MLE) but the profile likelihood method is much less time consuming.

When time-dependent covariates are included in the model, parameter estimation can be done without too much effort. It becomes slightly more complex and time-consuming when time-dependent covariates are included. Another limitation of the POM is that that little inference has been done to include censored data in the POM. Authors such as Murphy *et al.* (1997) and Cheng *et al.* (1995) addressed this limitation and presented POMs with censored data.

POM Advantages

According to Kirmani and Gupta (2001), the POM is a useful model when the probability of failure converges over time due the diminishing effect of covariate influence. Another advantage of the POM is that parameter estimation is not too complex when time-independent covariates are included in the model.

POM Disadvantages

Including time-dependent covariates increases the complexity and time consumption of parameter estimation. Also, little research has been done to include censored data in the POM model severely limiting the model in reliability.

Evaluation

In Table 3.2 the POM is evaluated according to each criteria with accompanying reasons for the given marks. Reasons presented are based on the content of the model presentation.

Table 3.2: Proportional Odds Model evaluation.

Criteria	**Mark**	**Reason**
Future potential	2	Limited potential in reliability
Implementation intricacy	3	Time-dependent covariates complexity
Reliability application	3	Limited application
Flexibility	3	Censored data modelling

3.3 Additive Hazards Model

Pijnenburg (1991) introduced the Additive Hazards Model (AHM) by modelling the reliability of the air conditioning systems in aircraft. The AHM can be given by:

$$h(x, z) = h_0(x) + \alpha \tag{3.3.1}$$

where $h_0(x)$ is the baseline FOM and ψ the functional term. Lin *et al.* (1998) indicated that α can be given by $\alpha = \lambda(\overline{z(x)} \cdot \overline{\gamma})$. The baseline FOM is only a function of time and is not affected by including covariates with α. Due to the fact that α is added to the baseline FOM, covariates have a additive effect rather than a multiplicative effect.

This property of the model enables α to be negative without necessarily resulting in a negative FOM. Gorjian *et al.* (2010) indicated that another useful property of AHM is that it is capable of modelling scenarios where the FOM is not zero at time zero. Both these properties are illustrated in Figure 3.2 where it is assumed that $\alpha > 0$.

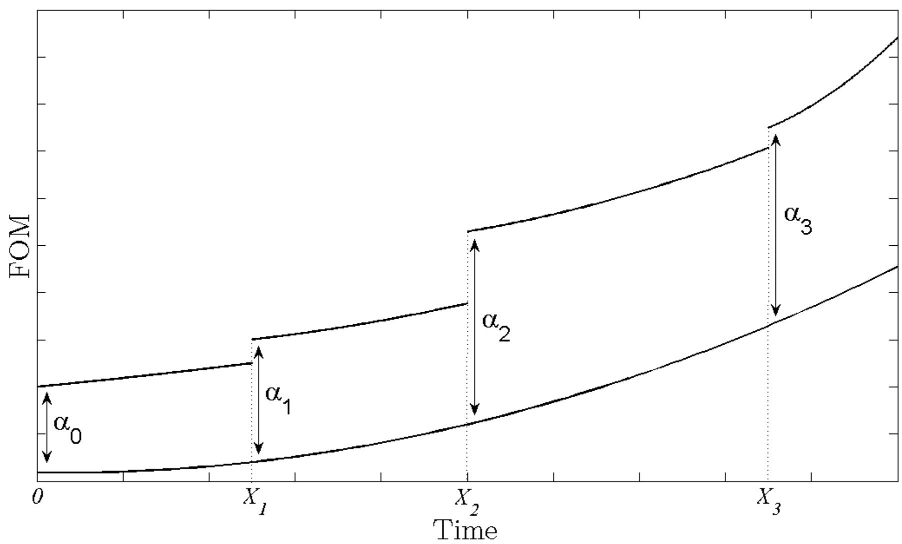

Figure 3.2: Graphical illustration of the AHM.

In Figure 3.2 the additive effect of the covariates are represented by α_i. It can also be seen that the FOM does not start at zero due to the addition of α_0 at time zero. Also, at each point of failure X_i, the FOM is seen to experience a 'jump' in FOM function value due to the assumption that $\alpha > 0$ and the additive affect of Equation 3.3.1. This 'jump' in function value indicates that some sort of maintenance action has been taken which influenced the FOM of the system.

As in the case of Figure 3.2 where $\alpha > 0$, the state of a system is in a WO state after maintenance action took place. On the other hand, scenarios where $\alpha = 0$, maintenance action had no effect on the FOM i.e., BAO. If however $\alpha \leq 0$, the system experienced a decrease in FOM after maintenance action took place i.e., BOWN. Note that in Chapter 2 it was stated that this thesis focusses on non-repairable systems which assumes that a system is in a GAN state after maintenance action has taken place. A non-repairable system can also be in a GAN state after maintenance action has taken place. However, although the system is in a GAN state, the system might be exposed to a higher FOM.

Parameter estimation can be done using the MLE and has proved to be a complex procedure. Pijnenburg (1991) confirmed this statement indicating the intricacy of the likelihood equation inference.

AHM Advantages

A key merit of the model is that it offers a non-zero FOM at time zero. In most situations the FOM won't be zero at time zero which renders this an useful property.

AHM Disadvantages

As mentioned by Lin *et al.* (1998), a limitation of the AHM is that a negative functional term is allowed. The result is a negative FOM estimate which is an unrealistic and unreasonable assumption.

Evaluation

Table 3.3 contains the evaluation of the AHM according to each criteria with an accompanying reason.

Table 3.3: Additive Hazards Model evaluation.

Criteria	**Mark**	**Reason**
Future potential	3	Good potential for repairable systems
Implementation intricacy	4	Popular estimation methods are used
Reliability application	4	Frequently applied in reliability
Flexibility	3	Limited non-repairable system applications

3.4 Proportional Hazards Model

According to Orbe *et al.* (2002), the Proportional Hazards Model (PHM) is the most commonly used model in failure data analysis. The PHM was developed by Cox (1972*a*) initially intended for the biomedical industry. Due to the success of the model in the biomedical industry, authors such as Ascher and Feingold (1984*a*) and Dale (1985) decided to apply the model in the field of reliability. The original model developed by Cox (1972a) is a product of an arbitrary and unspecified baseline FOM $h_0(x)$, and a functional term $\lambda(\overline{z(x)}\cdot\overline{\gamma})$. The FOM can be given by:

$$h(x,z) = h_0(x) \cdot \lambda(\overline{\overline{z(x)} \cdot \gamma}) \tag{3.4.1}$$

Before the discussion proceeds, it is necessary to define two assumptions of the PHM listed below.

1. Variables which affect the failure times of the items are included in the model.

2. For any two sets of time-independent sets of covariates, z_1 and z_2, for a single item, their associated FOM values are proportional with respect to time. Note that this assumption only holds for time-independent covariates.

The first assumptions states that all the environmental influences affecting the failure times of the item being studied are included in the model. Then, two sets of time-independent covariates are assumed to have proportional FOM functions. These two functions, $h_1(x,z)$ and $h_2(x,z)$, are also proportional to the baseline FOM $h_0(x)$. Figure 3.3 illustrates the proportional FOM assumption.

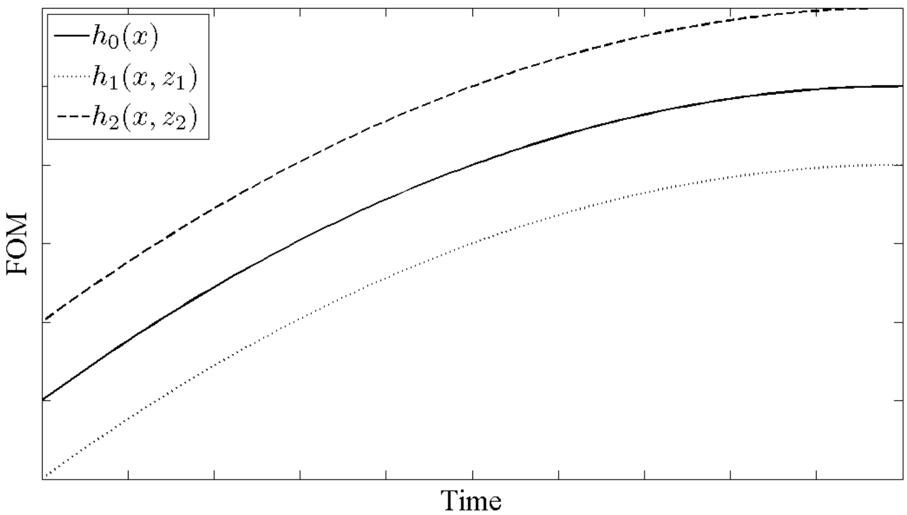

Figure 3.3: Graphical illustration of the proportional FOM assumption.

Furthermore, due to the construction of the model, covariates act multiplicatively on the FOM. The functional term, $\lambda(\overline{z(x)})$, can adopt several different forms. Some of these forms are: the logarithmic form, $\log(1 + \exp(\overline{z(x)} \cdot \overline{\gamma}))$; the inverse linear form, $1/(1 + (\overline{z(x)} \cdot \overline{\gamma})$; the linear from, $1 + (\overline{z(x)} \cdot \overline{\gamma}$ and the exponential form, $\exp((\overline{z(x)} \cdot \overline{\gamma})$. In most reliability applications, the exponential form is used whereby Equation 4.1.1 can be given by:

$$h(x,z) = h_0(x) \cdot \exp(\overline{z(x)} \cdot \overline{\gamma}) \tag{3.4.2}$$

Parameter estimation can be done using the Partial Likelihood Method (PLM) developed by Cox (1972*a*). Further developments of the model has indicated that the MLE can also be used to estimate the model parameters. Kalbfleisch and Prentice (1980) presents both the PLM and the MLE for the PHM.

A key advantage of this model is that no assumption has to be made about the baseline FOM of Equation 4.1.1. That means that parameter estimation can be done without having to make any assumption about the baseline FOM during model fitting procedures. A common form of the baseline FOM found in literature is the Weibull distribution. Substituting the Weibull distribution in Equation 3.4.2 yields:

$$h(x,z) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} \cdot \exp(\overline{z(x)} \cdot \overline{\gamma}) \tag{3.4.3}$$

where β and η are known as the shape and scale parameters respectively. This form of the PHM is known as the Weibull PHM and was presented by Jardine *et al.* (1987). Authors such as Ghasemi *et al.* (2009), Tian and Liao (2011) and Wong *et al.* (2010) applied the Weibull PHM in the field of reliability.

Seetharaman and Chintagunta (2003) indicates that alternative parametric distributions such as the inverse Gaussian, Erlang, Log-logistic, Exponential, Log-normal, Raleigh and Gompertz can be utilized for estimating the baseline FOM.

Another important property of the model is that it is possible to include censored events in the data. This enables modelling of items which have been taken out of service before having reached failure. An example of such a situation is where an item's dataset contains fewer failures than censored events. Such a study might indicate that maintenance took place too early and could have been postponed without encountering failure. Postponement might result in improved item performance and utilization.

Furthermore, a limitation of the model however is that small sample sizes should be avoided. Orbe *et al.* (2002) states that regression coefficients are biased towards certain covariates when small sample sizes are used.

PHM Advantages

The key merit of the PHM is that parameter estimation can be done without making an assumption about the baseline FOM. As with the POM model, covariates act multiplicative on the FOM which is a more reasonable and realistic assumption compared to models such as the AHM. A common form of the baseline FOM is the Weibull distribution. Application of this distribution is a special case of the PHM known as the Weibull PHM. The Weibull PHM is a flexible model and has repeatedly been applied in the field of reliability.

PHM Disadvantages

Using small sample sizes results in regression coefficients biased towards certain covariates.

Evaluation

Table 3.4 contains the evaluation of the PHM according to each criteria with an accompanying reason.

Table 3.4: Proportional Hazards Model evaluation.

Criteria	**Mark**	**Reason**
Future potential	5	Good future in reliability
Implementation intricacy	4	Popular estimation methods can be used
Reliability application	5	High application frequency in reliability
Flexibility	4	Weibull PHM exhibits good flexibility

3.5 Accelerated Failure Time Model

Pike (1966) introduced the Accelerated Failure Time Model (AFTM). The AFTM is a parametric model which estimates the relation between the log expected failure time and covariates in a familiar form of a linear regression equation. It allows covariates to act multiplicatively on the failure time of an item.

Cheung *et al.* (2001) showed that it is intuitive to analyze the effects of covariates using failure time ratios. In this section, the AFTM is presented in two ways. For the first representation, suppose that the AFTM estimates a log-linear regression lifetime $Y = \log(X)$. This is related to the covariate vector $\overline{(x)}$ via a linear model:

$$\log(X) = \overline{z(x)} \cdot \overline{\gamma} + \varepsilon \tag{3.5.1}$$

where ε is a measure of variability in the failure times. Substituting $\overline{\gamma}$ and $\overline{z(x)}$ into Equation 3.5.1 yields:

$$\log(X) = \overline{\gamma_0} + \overline{z_1} \cdot \overline{\gamma_1} + \overline{z_2} \cdot \overline{\gamma_2} + ... + \overline{z_m} \cdot \overline{\gamma_m} + \varepsilon. \tag{3.5.2}$$

If the coefficients of the above equation are exponentiated, $\exp(\overline{\gamma_m})$, it is known as the Time Ratio (TR). The TR can be used to estimate the failure time of an item. A $TR > 1$ implies that the time-to-event slows down, in other words increasing the time-to-event. On the other hand, a $TR < 1$ indicates that a shorter failure time is more likely. Therefore, covariates act multiplicatively on the failure time rather than on the FOM. Figure 3.4 illustrates different TR values and its effect of the failure probability time-scale of an item.

A common use of the AFTM is modelling electronic equipment to determine the failure probability of an item at different temperature levels over a given period of time. The AFTM also has repeatedly been applied in biomedical applications to model the effect different medication dosages may have.

The FOM of the AFTM can be written as:

$$h(x, \overline{z(x)}) = \lambda(\overline{z(x)}) \cdot h_0(\overline{z(x)} \cdot x) \tag{3.5.3}$$

where $h_0(\overline{z(x)} \cdot x)$ is the baseline FOM.

For the second representation, suppose that the reliability function is given by $R_X(x)$ and the accelerated reliability function by $\widehat{R_X(x)}$, resulting from stress or environmental influences. These influences are captured in covariate measurements. When covariates change it affects the model parameter values. Consequently, changes in covariates values results in time scale transformations of the

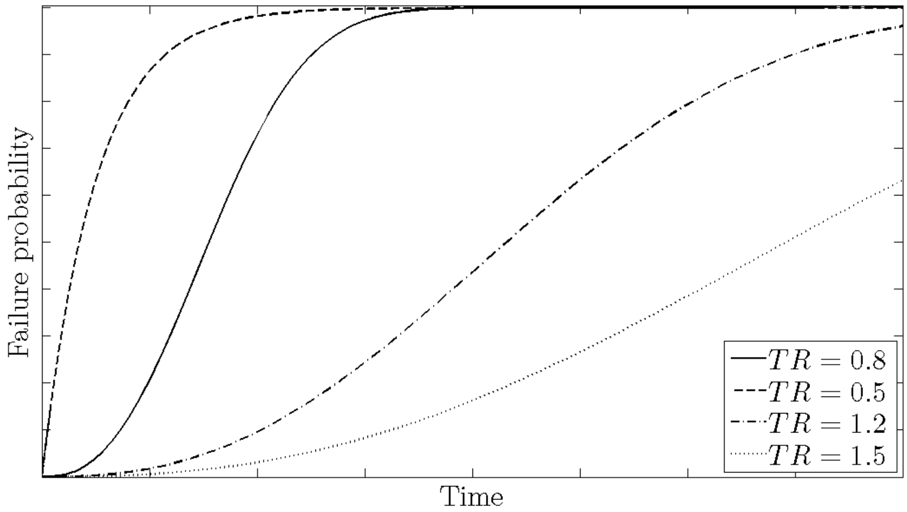

Figure 3.4: Different TR values and how it affects the failure probability time-scale of an item.

FOM and the other reliability equations. This was confirmed by Newby (1988) doing a study on fatigue crack growth applying both the AFTM and the PHM. For the AFTM, each of the reliability functions are shown to be related by:

$$\widehat{R_X(x)} = R[(x-c)/b] \tag{3.5.4}$$

where b is a scale parameter and c a location parameter. The AFTM model is similar to other regression models which assume that $(x-c)/b$ has a distribution related to a known parametric form. This leads to the use of shape, scale and location parameter families of distributions. With these distributions the failure times do not change with the level of covariate influence but is rather accelerated or decelerated.

Some probability density functions used in reliability analysis (which form part of the shape, scale an location families) are:

Weibull:

$$\widehat{f_X(x)} = \frac{k}{b} u^{k-1} \exp(-u^k), \quad u = x/b \tag{3.5.5}$$

Gamma:

$$\widehat{f_X(x)} = \frac{1}{b\Gamma(k)} u^{k-1} \exp(-u)], \quad u = x/b \tag{3.5.6}$$

Log-normal:

$$\widehat{f_X(x)} = \frac{1}{buk\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(u)}{k}\right]^2\right\} u^{k-1} \exp(-u)], \quad u = (x-c)/b \tag{3.5.7}$$

Inverse Gaussian:

$$\widehat{f_X(x)} = \frac{1}{b}\sqrt{\frac{k}{2\pi}} u^{-3/2} \exp\left\{-\frac{-k(u-1)^2}{2u}\right\}, \quad u = x/b \tag{3.5.8}$$

where b and c are as defined and k a shape parameter. If it is possible to find a distribution which adequately describes the data, a new definition of the probability density function is $f_X(\underline{x,k})$ with a corresponding AFTM density function of $f_X((x/b_i),k)$, where $b_i = b(\underline{z(x)},q)$ and q a coefficient parameter.

Quasi-linear models are assumed to be in the form of $b = k(\overline{z(x)} \cdot c)$ and $b = k(\overline{z(x)} \cdot a)$. Other forms of b are:

1. Constant.
2. Linear function of stress, $b = (\overline{z(x)} \cdot a)$.
3. Exponential function, $b_i = \exp(z_i \cdot a)$.
4. Inverse exponential Arrhenius model, $b_i = \exp(-1/(\overline{z(x)} \cdot a))$.

Parameter estimation can be done using either the partial likelihood method developed by Cox (1972a) and the MLE. Supporting this statement is Newby (1988), Shyur *et al.* (1999), Shyur *et al.* (1999) and Orbe *et al.* (2002). These methods are often used in literature which reduces the implementation complexity thereof. Qi (2009) used this method by maximizing the log-likelihood function to estimate the model parameters.

Application in the field of reliability has however been limited. A reason for this might be that parameter estimation is done by assuming a distribution for the failure times. Failure time distributions are intricate to identify and time consuming to determine. Another reason might be the fact that the inclusion of censored data has not seen much development.

AFTM Advantages

The model directly models the influence covariates have on the failure time by accelerating or decelerating it by a constant value making it much easier to interpret the results. Parameter estimation is done using familiar parameter estimation methods.

AFTM Disadvantages

Although having previously been applied in the field of reliability, the AFTM is relatively unfamiliar and rarely used. Reasons might include unknown failure time distribution assumption, parameter estimation is time consuming and little research being done to include censored events.

Evaluation

Table 3.5 contains the evaluation of the AFTM according to each criteria with an accompanying reason.

Table 3.5: Accelerated Failure Time Model evaluation.

Criteria	**Mark**	**Reason**
Future potential	3	Some potential for reliability
Implementation intricacy	4	Popular estimation methods are used
Reliability application	3	Few number of reliability applications
Flexibility	3	Limitation on censored data inclusion

3.6 Proportional Intensity Models

The Proportional Intensity Model (PIM) was introduced by Cox (1972*b*). It is a product of a baseline ROCOF and a functional term. The functional term is a function of covariates and may or may not be a function of time.

The construction of the PIM is quite similar to that of the PHM. However, Goldman (1967) states that the main difference between the two models are that the PIMs failure mechanism follows a stochastic point process.

A stochastic point processes is a countable aggregate of failure times stochastically distributed in some Cartesian space, i.e. R. This is in contrast to the PHM where a failure mechanism is assumed to follow a known distribution such as the

Weibull distribution. As a result, it is often found that the PIM does not assume system renewal at failure.

Now, consider a baseline ROCOF $\iota_{U_0}(t)$, and the functional term $\lambda(t, z(t))$. The PIM model can then be given by:

$$\iota_U(t) = \iota_{U_0}(t) \cdot \lambda(\overline{z(t)}, t) \tag{3.6.1}$$

Possible forms of the baseline ROCOF, $\iota_{U_0}(t)$, can be adopted. Three of these forms are: power-law, αt^{β}; constant α and log-linear, $\alpha\beta^{t}$. A form adopted in the field of reliability for the functional term, $\lambda(t, z(t))$, is given by $\exp(\overline{z(t)}, \overline{\gamma})$. Estimation of the $\overline{\gamma}$ parameters is done using the MLE. Due to the large amount unknown parameters, parameter estimation is a complex process.

Authors such as Love and Guo (1991), Percy and Alkali (2006), Jiang *et al.* (2006), Lugtigheid *et al.* (2007) and Percy and Alkali (2007) have illustrated how the PIM can be implemented in the field of reliability.

PIM Advantages

Parameter estimation can be done without making any assumptions about the baseline ROCOF. Also, the PIM does not assume that a system is renewed at failure such as the PHM.

PIM Disadvantages

The PIM is limited to repairable systems due to the fact that it does not assume system renewal at failure. However, Vlok *et al.* (2004) introduced a variation of the PIM for both repairable and non-repairable systems. Lastly, the process of parameter estimation is complex due to the large number of unknown parameters.

Evaluation

Table 3.6 contains the evaluation of the PIM according to each criteria with an accompanying reason.

3.7 Extended Hazard Regression Model

The Extended Hazard Regression Model (EHRM) was introduced by Ciampi and Etezadi-Amoli (1985) and Etezadi-Amoli and Ciampi (1987) and includes both the PHM and the AFTM. A general expression for the FOM is given by:

Table 3.6: Proportional Intensity Model evaluation.

Criteria	**Mark**	**Reason**
Future potential	4	Good future potential for reliability
Implementation intricacy	3	Complex parameter estimation
Reliability application	4	Often applied in reliability
Flexibility	2	Some limitation to system type

$$h(x, z) = g_1(\overline{z(x)} \cdot \overline{a}) \cdot h_0[g_2(\overline{z(x)} \cdot \overline{b}) \cdot x] \tag{3.7.1}$$

Where $\overline{a}$ and $\overline{b}$ are regression coefficient vectors and $g_1(\cdot) = g_2(\cdot) = 1$ at time zero. When $b = 0$, Equation 3.7.1 reduces to the PHM and when $a = b$, Equation 3.7.1 reduces to the AFTM. Therefore indicating the presence of properties of both the PHM and AFTM in the EHRM. The multiplicative effect of the covariates on the FOM of the PHM is interpreted by the $\overline{b}$ and the time-scale changing effect of the AFTM is interpreted by the a.

Estimation of the baseline FOM can be done using the method of quadratic splines. A quadratic spline is a piecewise polynomial of n degrees with pieces joining at a number locations called *knots*, failure times. By illustration, Figure 3.5 illustrates how quadratic splines can be used to join *knots* to estimate a baseline FOM.

It should however be noted, utilizing quadratic splines to estimate the baseline FOM does not ensure a positive FOM. For example, if the 5^{th} failure in Figure 3.5 were close to zero, the spline arc could have intercepted the y-axis resulting in a negative FOM.

Furthermore, application of the model is fairly complex due to the fact that the unknown parameters, β and η, have to be estimated simultaneously. Elsayed *et al.* (2006) sets out a method to estimate these parameters using the MLE but also comments on the complexity involved.

EHRM Advantages

The MLE is used to estimate the unknown model parameters of the EHRM. Shyur *et al.* (1999) states that the EHRM is a useful regression model with the future potential in the field of reliability.

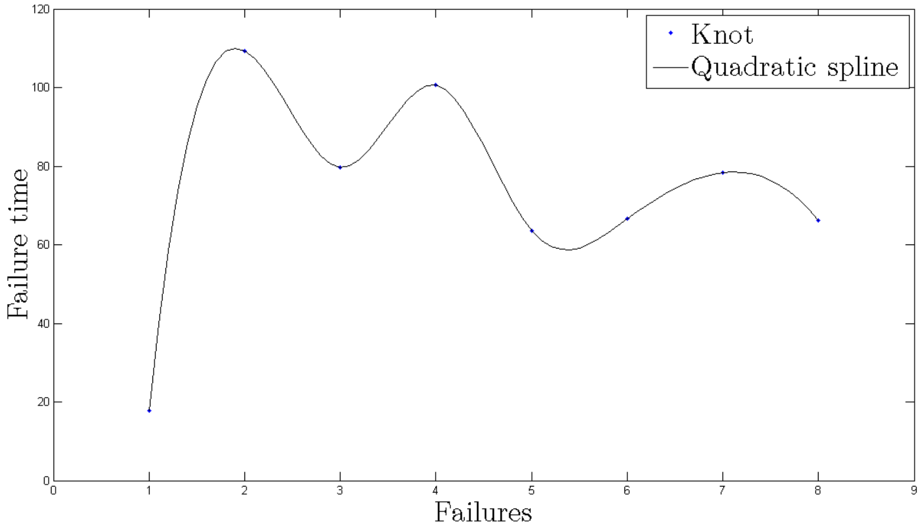

Figure 3.5: Baseline Foce Of Mortality estimation using quadratic splines.

EHRM Disadvantages

The EHRM has a major drawback where a positive FOM cannot be ensured placing a great deal of uncertainty in the model's ability to model a reliability situation realistically.

Evaluation

Table 3.7 contains the evaluation of the EHRM according to each criteria with an accompanying reason.

Table 3.7: Extended Hazard Regression Model evaluation.

Criteria	**Mark**	**Reason**
Future potential	2	Limited reliability application
Implementation intricacy	2	Complex parameter estimation
Reliability application	3	Seldom application
Flexibility	3	Baseline FOM estimation

3.8 Model Selection

In Chapter 2 a primary selection process was done to identify models with the potential achieve the thesis objectives. In this chapter each of those selected models were presented and evaluated according to criterion given in Section 3.1.

Models were evaluated according to each criteria and given a mark out of five. In Table 3.8 each model's total mark is calculated taking into account the weights of the criterion.

Table 3.8: Selection of most suitable regression model.

Regression Model Evaluation							
Criterion	Weight	POM	AHM	PHM	AFTM	PIM	EHRM
Future potential	2^1	2	3	5	3	4	2
Reliability application	2^2	3	4	4	4	3	2
Implementation intricacy	2^3	3	4	5	3	4	3
Flexibility	2^4	3	3	4	3	2	3
	Total Score:	31	98	126	90	81	82

This indicates that the PHM is the most suitable regression model for the thesis. From this point forward, the sole focus of the thesis is on the PHM. Chapter 4 provides an in-depth study of the PHM with a keen focus on the mathematical model, parameter estimation, model fitting procedures and a decision model.

3.9 Conclusion

This chapter addressed the second item of the methodology whereby models identified in Chapter 2 are evaluated according to four criterion reflecting the applicability of the model to achieve the objectives of this research project. The most appropriate model was then selected based on these criterion and it was found that the PHM is the most suitable model for this thesis.

Chapter 4

Proportional Hazards Model

4.1 Introduction

Several regression models were evaluated in Chapter 3 of which the Proportional Hazard Model (PHM) was chosen as the most suitable model for this thesis. In Chapter 4, the methodology is to perform a comprehensive study on the PHM focussing on the mathematics and practical implementation thereof. The chapter concludes with a brief look at some decision models which are used together with the PHM to develop ACPs.

When the PHM was introduced by Cox (1972*a*), its initial intention was for it to be used for biomedical applications. After the success of the model in this field, several authors discovered that the PHM can be used in reliability applications. Consequently, an abundance of reliability applications can be found in literature. Some examples of where the model has been applied in reliability applications are: mining haul truck wheel motors (Jardine *et al.*, 2001), circulating pumps (Vlok *et al.*, 2002), locomotive wheel set bearings (Feng *et al.*, 2009), preventive maintenance scheduling (Samrout *et al.*, 2009), military vehicle diesel engines (Wong *et al.*, 2010), electrical transformers (Wu and Ryan, 2011), solder joint's life span (You *et al.*, 2011) and semiconductor manufacturing (Pampuri *et al.*, 2011).

These examples utilize different variations of the model to compensate for the specific needs of each application. For this reason, it is not practical to discuss every variation of the model. The discussion is limited to the mathematical theory necessary to apply the model in practical situations.

The PHM model is a product of an arbitrary and unspecified baseline FOM $h_0(x)$, and a functional term $\lambda(x \cdot \overline{z(x)})$. The functional term is a function of time and covariates. Note, if the covariates are independent of time, then the functional term is a function of covariates only, i.e. $\lambda(\overline{z})$.

Let $h(x, \overline{z})$ represent the FOM at time x with corresponding covariate vector $\overline{z_i(x)} = [z_1, z_2, z_3, ..., z_n]$. For the sake of generality, theory presented in this thesis consider time-dependent covariates. The FOM is therefore given as:

$$h(x, \overline{z(x)}) = h_0(x)\lambda(x \cdot \overline{z(x)}) \tag{4.1.1}$$

In Chapter 3 it was indicated that several forms of the functional term are considered and that most reliability applications of the PHM adopt the exponential form, i.e. $\exp(\overline{z(x)}, \overline{\gamma})$. Substituting this form of the functional term in Equation 4.1.1 yields:

$$h(x, \overline{z(x)}) = h_0(x) \exp(\overline{z(x)} \cdot \overline{\gamma}) \tag{4.1.2}$$

where $\overline{\gamma}$ is a row vector containing the regression coefficients. These coefficients are estimated during model fitting procedures discussed in Section 4.5.

Two forms of the PHM are discussed in this thesis. The first form is the semi-parametric form which does not make any assumptions about the baseline FOM. The second form fully specifies the baseline FOM and is known as the fully-parametric form. Then, an algorithm is discussed which was used for numerical model fitting procedures. And finally, decision models are introduced with which improved maintenance plans can be developed.

4.2 Censoring

In Chapter 2 censoring was briefly discussed where it was stated that censoring plays a crucial role in reliability analysis because it allows the inclusion of items in the model which have not yet failed. In order to determine model parameters, the likelihood function needs to be derived whereby it is necessary to consider the censoring mechanism acting on the failure data.

Likelihood is the conditional probability that an item, which has already failed, yields a certain outcome. This is in contrast to probability which is based on events that have not yet occurred.

Let n items be put on test where X_i indicates failure times and $\overline{z_i(x)}$ corresponding covariate vectors. A subjective variable, c_i, is introduced indicating either an observed failure, $c_i = 1$, or a censored observation, $c_i = 0$. In the reliability field, three censoring types are observed:

1. *Drop Out*: The item fails in the duration of an experiment resulting in an uncensored observation.
2. *Termination of Experiment*: The item has not yet failed at the end of an experiment resulting in a censored observation.
3. *Follow-up*: The item is lost during the duration of an experiment resulting in a censored observation.

Figure 4.1 illustrates three failure times to indicate each form of censoring.The first item failed at time X_1 resulting in an uncensored observation. Item number two hadn't failed at the end of the experiment resulting in a censored observation X_2^+. The third item was lost during the experiment which resulted in another censored observation X_3^+.

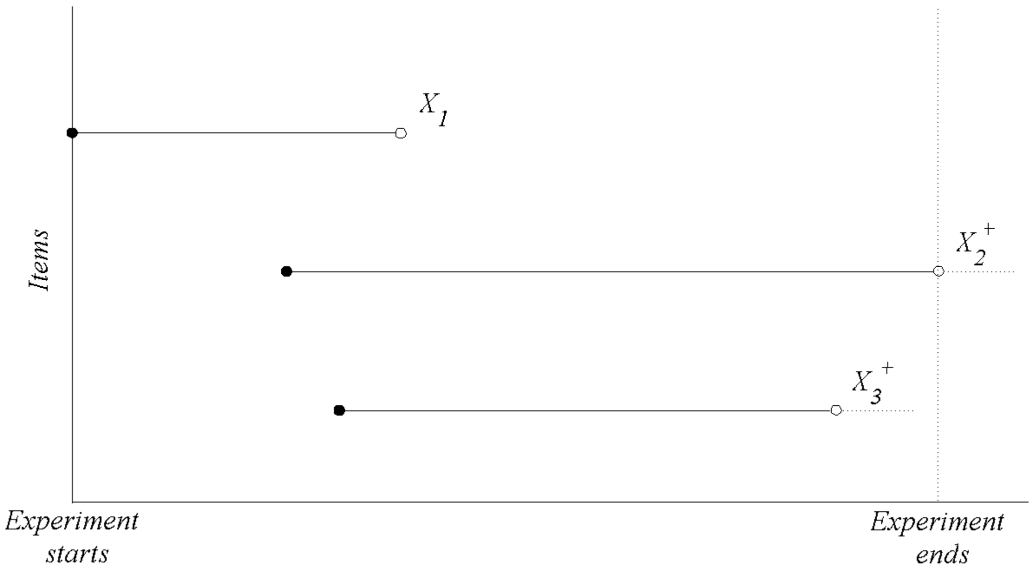

Figure 4.1: Three types of Random censoring.

Three mechanisms of censoring are discussed in this chapter, namely (a) random censoring, (b) type I censoring and (c) type II censoring.

Random censoring

Let the time to censoring be C_i with probability density and survivor functions $g_i(x)$ and $G_i(x)$ respectively. Also $C_i = C_1, C_2, C_3, ..., C_n$ are randomly and independently distributed of each other and of the failure times $X_i = [X_1, X_2, X_3..., X_n]$.

It might seem confusing that C_i and X_i are similar. To clear up confusion, the i^{th} individual may or may not be observed due to censoring and therefore failure times, X_i, and censoring times, C_i, are not the same where censored observations are found.

Using random censoring can be used to include type I censoring. Random censoring also allows items to enter the study at random instances. Kalbfleisch and Prentice (1980) indicated that for random censorship the probability of failure time X_i occurring in $(x, x + dx)$ is given by:

$$P\left[X_i \in (x, x + dx), c_i = 1; \overline{z_i(x)}, \overline{\gamma}\right] = G_i(x + 0)f(x; \overline{z_i(x)}, \overline{\gamma})dt \tag{4.2.1}$$

and

$$P\left[X_i \in (x, x + dx), c_i = 0; \overline{z_i(x)}, \overline{\gamma}\right] = g_i(x)F(x + 0; \overline{z_i(x)}, \overline{\gamma})dt \tag{4.2.2}$$

where if F is continuous, $F(x + 0; \overline{z_i(x)}, \overline{\gamma}) = F(x; \overline{z_i(x)}, \overline{\gamma})$.

Type I censoring

Here censoring times of an item is fixed at the start of an experiment wherby a maintenance eam knows exactly when maintenance is going to be performed on an item. Typically this is a maintenance strategy such as Usage Based Maintenance.

If it is found that $G_i(x)$ do not involve any parameters of interest, i.e. $\overline{\gamma}$, Equations 4.2.1 and 4.2.2 simplify to:

$$L(\overline{\gamma}) = \prod_{i=1}^{n}\left[f(X_i; \overline{z(x)}; \overline{\gamma})^{c_i} \cdot F(X_i; \overline{z(x)}; \overline{\gamma})^{1-c_i}\right] \tag{4.2.3}$$

which is the likelihood function for type I censoring. Some equations, such as Equation 4.2.3, are introduced in this section although their use shall become clear in sections to follow. It is the product of each item's likelihood function given by:

$$L(\bar{\gamma}) = \prod L_i(\bar{\gamma}) \tag{4.2.4}$$

where $L_i(\overline{\gamma})$ is $f(X_i, \overline{z(x)}, \overline{\gamma})$ for a failure and $F(X_i, \overline{z(x)}, \bar{\gamma})$ for a censored observation. Type I censoring can typically be found in Use Based Maintenance (UBM) situations.

Type II censoring

Type II censoring times are unknown at the beginning of an experiment. Censoring time is a function of the amount of items which has failed. Items are operated until r out of n items have failed. So the r shortest failure times are observed to fail and the remaining $n-r$ items are censored.

Hence, data consist of the r shortest failure times and $n-r$ censored observations. The likelihood function can be given by:

$$L = \frac{n!}{(n-r)!}(f(X_1)\cdots f(X_r))(F(X_r))^{n-r} \tag{4.2.5}$$

For example, if a plant requires r out of n pumps operational for acceptable and reliable operation, type II censoring can be used indicate when downtime is necessary to perform maintenance in order to ensure that r pumps are available.

For both Type I and Type II censoring, the likelihood construction is the same. In both cases items that fail contribute to the $f(X_i)$ term and censored items contribute to the $F(X_i)$ term.

4.3 Semi-parametric PHM

The main advantage of this form is that baseline FOM is left unspecified. This model utilizes the FOM in Equation 4.1.2 with a functional term in the exponential form, i.e. `exp`$(\overline{z(x)} \cdot \overline{\gamma})$, which is given as:

$$\exp(\bar{z}(x) \cdot \bar{\gamma}) = \exp\left[\sum_{m=1}^{k} z_m(x) \cdot \gamma_m\right] \tag{4.3.1}$$

where k indicates the number of covariate measurements. From this it can be seen that the covariates are fully specified whereas the baseline FOM is unspecified. For this reason the model is called the semi-parametric model.

Semi-parametric PHM regression coefficient inference can be done using a method introduced by Cox (1972a) known as the method of *partial likelihood.* As an introduction to this estimation method, *likelihood* and the *Maximum Likelihood Estimate* are discussed.

4.3.1 Maximum Likelihood Estimate

In Chapter 3 the method of the Maximum Likelihood Estimate (MLE) was referred to on multiple occasions. Most of these regression models utilize some form of this method to determine its' regression coefficients. It's important to note that the MLE is suitable for both the semi-parametric and fully-parametric PHM parameter estimation. It is therefore necessary to briefly look at this method before the methods of *partial likelihood* and *fully-parametric* PHM are introduced.

To illustrate the functional purpose of the MLE, consider a *random censoring* mechanism and that the censoring times have no influence on the failure times. From Equation 4.2.3, the log-likelihood for the maximization process can be given by:

$$L(\bar{\gamma}) = \prod_{i=1}^{n} \left[f(X_i;\bar{\gamma})^{c_i} \cdot F(X_i;\bar{\gamma})^{1-c_i} \right] = \sum_{a} \log f(X_i;\bar{\gamma}) + \sum_{b} \log F(X_i;\bar{\gamma}) \quad (4.3.2)$$

where a and b represent sums of the total amount of failures and censored observations, respectively. The value of $\overline{\gamma}$ that maximizes Equation 4.3.2 is the most suitable since it maximizes the probability of failure occurrence in the observed data set. The log-likelihood function is maximized due to the fact that that it is more convenient to maximize the log-likelihood function rather than the likelihood function. This process is peformed by Orbe *et al.* (2002), Vlok *et al.* (2002) and Ghasemi *et al.* (2009).

Several software packages and numerical methods are available with which the maximum likelihood function can be estimates. Carstens *et al.* (2011) developed a metaheuristic algorithm for this thesis to estimate the PHM regression coefficients and discussed later in Chapter 4.

4.3.2 Marginal Likelihood

Before the method of *partial likelihood* is discussed, a brief overview of the marginal likelihood is given. The method of marginal likelihood can be used to estimate the parameters of a semi-parametric PHM with unknown model parameters and unspecified baseline FOM function.

Suppose n items are put on test and are observed to fail at failure times X_i, with corresponding covariates z_i. Order statistic is introduced as $O(x) = [X_1, X_2, X_3, ..., X_n]$ where the $X_i's$ are ordered from smallest to largest. The rank statistic $r(x) = [1, 2, 3, ..., n]$ is also introduced where the i notation refers to the

i^{th} order statistic.

For example, let the observed failure times of five items be $X_1 = 45, X_2 = 65, X_3 = 12, X_4 = 98, X_5 = 72$. As a result, the order statistic is given by $O = [12, 45, 65, 72, 98]$ and the rank statistic by $r(x) = [3, 1, 2, 5, 4)]$.

Although the order statistic contains the failure times, the rank statistic carries the only information about the regression coefficients when the baseline FOM is unspecified. According to Barnard (1963), the rank statistic is sufficient for regression coefficient inference when working with an unspecified baseline FOM. As a result, the Marginal Likelihood (ML) is proportional to the probability that the rank statistic is observed which is given by:

$$\frac{\exp(\sum_{i=1}^{n} (z_i \cdot \bar{\gamma}))}{\prod_1^n \left[\sum_{l \in R_{(X_i)}} \exp(z_i \cdot \bar{\gamma})\right]} \tag{4.3.3}$$

where $R(X_i)$ is a set of items at risk prior to X_i where $R(X_i) = i, (i+1), ..., (n)$. To accommodate censoring, it is necessary to modify Equation 4.3.3. For example, let the observed failure times of four items be $78^*, 34, 98^*, 49$, with the asterisk indicating censored observations. Consequently, the rank statistic is known to be one of the following six possibilities in sample set given by:

$$\begin{array}{ccc} [3,2,4,1]; & [3,4,2,1]; & [3,2,1,4]; \\ [3,4,1,2]; & [3,1,2,4]; & [3,1,4,2]. \end{array}$$

In order to estimate the regression coefficients while incorporating censored observation, the ML that the rank statistic should be one of these possibilities can be used. Kalbfleisch and Prentice (1980) states that it's reasonable to assume that exact censoring times do not contribute to the inference of the regression coefficients.

Let k items give rise to failure times of $X_1 < X_2 < X_3 < ... < X_k$ with corresponding covariates $z_i = [z_1, z_2, z_3, ..., z_k]$. It can the be shown that the ML that one of the ranks in the sample set should occur is given by:

$$\frac{\exp(\sum_{i=1}^{k} (z_i \cdot \bar{\gamma}))}{\prod_{i=1}^k \left[\sum_{l \in R(X_i)} \exp(z_i \cdot \bar{\gamma})\right]} \tag{4.3.4}$$

The same approach can be used to accommodate tied failure times. Tied failures are two or more failure times of equal length. Let d of n items be observed to fail at failures times $X_i = [X_1, X_2, X_3, ..., X_k]$ with $X_1 < ... < X_k$ and $\sum di =$

n. Peto (1972) and Breslow (1975) state that the regression coefficients are well estimated using the likelihood function given by:

$$L = \prod_{i=1}^{k} \frac{\exp(\overline{s_i(x)} \cdot \bar{\gamma})}{\left[\sum_{l \in R(X_i)} \exp(\overline{z_l(x)} \cdot \bar{\gamma}) \right]^{d_i}} \tag{4.3.5}$$

where $s_i = \sum z_{ij}$ is the sum of the covariates of the items observed to fail at X_i. The value of $\overline{\gamma}$ that maximizes Equation 4.3.5 is the most suitable for the given data set.

Unfortunately the method of ML has not seen much application. However, Cox (1975) showed how the ML is used to develop the PL. For this reason the method of ML was discussed here and it is shown in Section 4.3.3 how similar the two methods are.

4.3.3 Partial Likelihood

In Section 4.3.2 the ML was introduced which can be used to estimate regression coefficients in the presence of an unspecified baseline FOM. The method of Partial Likelihood (PL) was introduced by Cox (1975) to estimate regression coefficients without regard to the baseline FOM.

Therneau and Grambsch (2000) state that the PL is not a likelihood in the sense that it is the probability of observing a dataset which is proportional to the marginal or conditional probability. The method can nonetheless be used as a likelihood for regression coefficient inference. However, Cox state that the PL contains sufficient information about the regression coefficients for data with censored observations.

Suppose n items are observed to fail at $X_i = [X_1, X_2, X_3, ..., X_n]$ with corresponding covariates $z_i = [z_1, z_2, z_3, \underline{..., z_n}]$. The baseline FOM is obtained from the probability density function, $f(X_i, \underline{z_i(x)}, \bar{\gamma}, h_0(x))$. The failure times are transformed in a one-on-one manner using auxiliary variables such that $A_1, B_1, ..., A_k, B_k$. Then, let $A^k = (A_1, ..., A_k)$ and $B^k = (B_1, ..., B_k)$ be such that the joint density of $A^{(n)}$ and $B^{(n)}$ is given by:

$$\prod_{k=1}^{n} f(b_k | \, b^{(k-1)}, a^{(k-1)}; \bar{\gamma}; h_o(x)) \cdot \prod_{k=1}^{n} f(a_k | \, a^{(k)}, a^{(k-1)}; \bar{\gamma}) \tag{4.3.6}$$

Here the term on the right hand side is known as the partial likelihood of $\overline{\gamma}$ and *partial* since it only is part of Equation 4.3.6. It can also be noted that the partial likelihood does not contain the baseline FOM.

Next, consider the set $R(X_i)$ of individuals at risk prior to X_i. The PL that item i fails at X_i, given that the set of items $R(X_i)$ are at risk, is given by:

$$\frac{h(X_i, \overline{z_i(x)})}{\sum_{l \in R(X_i))} h(X_i, \overline{z_i(x)})} = \frac{\exp(\overline{z_i(x)} \cdot \bar{\gamma})}{\sum_{l \in R(X_i)} \exp(\overline{z_i(x)} \cdot \bar{\gamma})} \tag{4.3.7}$$

where $i = 1, 2, 3, ..., k$. Once again notice that the regression coefficients can be estimated in the absence of the baseline FOM. Recall the likelihood in Section 4.2 given by:

$$L(\bar{\gamma}) = \prod L_i(\bar{\gamma}) \tag{4.3.8}$$

indicating that likelihood is the product of each individual item's likelihood. If the right hand side of Equation 4.3.7 is taken to be an individual's PL, the PL is constructed by substituting this value into Equation 4.3.8 which yields:

$$L = \prod_{i=1}^{k} \frac{\exp(\overline{z_{(i)}(x)} \cdot \bar{\gamma})}{\sum\limits_{l \in R(X_i)} \exp(\overline{z_{(i)}(x)} \cdot \bar{\gamma})} \tag{4.3.9}$$

which is identical to the ML in Equation 4.3.4. To accommodate tied failure times, the same equation that was derived for ML, Equation 4.3.5, can be used in the case of the method of PL. As in the case of the method of ML, the regression coefficient vector $\overline{\gamma}$ that maximizes Equation 4.3.8 or Equation 4.3.9 is the most suitable for the given data set.

4.4 Fully-parametric PHM

The semi-parametric form of the PHM can be fully parametrized by specifying the form of the baseline FOM. A suitable form of the baseline FOM used in the reliably field is the Weibull distribution due to its flexibility. As a result, this form of the model is called the Weibull PHM.

4.4.1 Model

The FOM of the fully-parametric Weibull PHM is given by:

$$h(x, \overline{z(x)}) = \left(\frac{\beta}{\eta}\right)\left(\frac{x}{\eta}\right)^{\beta-1} \exp\left(\overline{z_i(x)} \cdot \bar{\gamma}\right) \tag{4.4.1}$$

where $\beta > 0$ and $\eta > 0$ are the shape and scale parameters, respectively. As stated in Chapter 2, if one of the reliability functions are known, inference about the remaining three equations is possible, i.e. $f(x, \overline{z(x)})$, $F(x, \overline{z(x)})$ and $R(x, \overline{z(x)})$. The relation between the FOM function and the reliability function is repeated here for convenience. According to Vlok (1999), the reliability function is given by:

$$R(X_i) = \exp\left(-\int_o^x h(X_i)dx\right) = \exp\left(-(X_i/\eta)^\beta\right) \tag{4.4.2}$$

where if a variable $P_i = (X_i/\eta)^\beta$ is considered, where P_i has a unit negative exponential distribution. Consider an item at time X_i under the influence of time-independent covariates. According to Equations 6.1.1 and 4.4.2, the reliability of an item can be estimated with the following expression:

$$R(x, \bar{z}) \;= \exp\;\left(-\int_0^{X_i}\left(\frac{\beta}{\eta}\right)\left(\frac{\mathrm{x}}{\eta}\right)^\beta dx \exp(\overline{z(x)}\cdot\bar{\gamma})\;\right) \tag{4.4.3}$$

$$= \exp\left(-(X_i/\eta)^\beta \exp(\overline{z(x)}\cdot\bar{\gamma})\right) \tag{4.4.4}$$

Where $P_i = (\mathrm{X}_i/\eta)^\beta \exp(\overline{\mathrm{z(x)}}\cdot\bar{\gamma})$, again with a unit negative exponential distribution. The next step is to consider time-dependent covariates whereby the reliability is given by:

$$R(x, \overline{z(x)}) \;\&\; = \exp\left(-\int_0^{X_i}\left(\frac{\beta}{\eta}\right)\left(\frac{\mathrm{x}}{\eta}\right)^\beta \exp(\overline{z(x)}\cdot\bar{\gamma})dx\right)\;\& \tag{4.4.5}$$

$$= \exp\left(-\int_0^{X_i} \exp(\overline{z(x)}\cdot\bar{\gamma})d((x/\eta)^\beta)\right) \tag{4.4.6}$$

where $P_i = \int_0^{X_i} \exp(\overline{z(x)}\cdot\bar{\gamma})d((x/\eta)^\beta)$ and also has a unit negative exponential distribution.

4.4.2 Regression Coefficient Estimation

Three techniques were discussed which is used to estimate the regression coefficients of some form of the PHM, i.e. maximum likelihood estimate, marginal likelihood and partial likelihood. Regression coefficients of the semi-parametric PHM is estimated using the methods of ML and PL. In the case of the fully-parametric PHM, the method of MLE is used to estimate the regression coefficients using the full likelihood function given by:

$$L(\beta, \eta, \overline{\gamma}) = \prod_i h(X_i, \overline{z_i(X_i)}) \prod_j R(X_j, \overline{z_i(X_j)}) \tag{4.4.7}$$

In Equation 4.4.7, i indexes failure observations and j indexes failure observations and censored observations, i.e. $j = 1, 2, 3, ..., n$. Substituting Equations 4.4.6 and 6.1.1 into Equation 4.4.7 yields:

$$\begin{aligned} L(\beta, \eta, \overline{\gamma}) = &\prod_i \left(\frac{\beta}{\eta}\right)\left(\frac{X_i}{\eta}\right)^{\beta} \exp(\overline{z(x)} \cdot \overline{\gamma}) \cdot \\ &\prod_j \exp\left[-\int_0^{X_i} \exp(\overline{z(x)} \cdot \overline{\gamma}) d((x/\eta)^{\beta})\right] \end{aligned} \tag{4.4.8}$$

However, this form of the full likelihood requires some complex computational effort. Therefore it is preferred to manipulate this form to the log-likelihood form, $l(\beta, \eta, \overline{z(x)})$. Maximization of the log-likelihood yields equivalent model parameters $(\beta, \eta, \overline{z(x)})$ as the full-likelihood and can for this reason be used instead. The log-likelihood is given by:

$$\begin{aligned} L(\beta, \eta, \overline{\gamma}) = \ &\Phi \ln(\beta/\mu) + \sum_i \ln\left[(X_i/\eta)^{\beta-1}\right] + \sum_i (\overline{z(x)} \cdot \overline{\gamma}) - \\ &\sum_i \int_0^{X_i} \exp(\overline{z(x)} \cdot \bar{\gamma}) d((x/\eta)^{\beta}) \end{aligned} \tag{4.4.9}$$

where Φ indicates the number of failure observations.

A methaheuristic algorithm was used for the maximization of Equation 4.4.9. The development and evaluation of this algorithm is discussed next. Note, from this point forward all efforts are aimed at the fully-parametric model and its' application. The rest of this chapter deals with its' application and Chapter 5 presents a case study in which the fully-parametric PHM is applied in a practical situation.

4.5 Model Fitting Procedure Of the Fully-Parametric Proportional Hazards Model

Several methods exist to solve non-linear problems such as the maximization of Equation 4.4.9. Some of these methods include quadratic programming, fractional programming, non-linear programming, stochastic programming, calculus of variations, metaheuristic, Nelder-Mead, Snyman's trajectory and Newton-Raphson.

Most of these models have been applied successfully in a number of different industries. Although these models have been proven in industry, some of them are developed in such a way that it is not applicable in the reliability field. Three main problems were found with these models:

- Initial value requirement.
- Large parameter spaces.
- Inconsistent solution convergence.

Most of the given models require initial values in order to determine a near optimal solution. In the case of the fully-parametric PHM, the initial values refer to the β, η and $\overline{\gamma}$ values. In most cases the parameter values are totally unknown. This leads to the second problem where parameters have to be found in large search spaces.

In most reliability situations, the parameters search space of each parameter is quite large due to the fact that the answer is unknown. This complicates initial value selection due to the second problem. Some models have shown to converge to different solutions if different initial values are used. From this it is clear that a model needs to be selected which does not have these problems.

Models without these problems are metaheuristic algorithms. These algorithms are used to find solutions to problems with large parameter search spaces without the need of initial values. This indicated that metaheuristic algorithms might be suitable for this thesis and had to be investigated.

Metaheuristic algorithms find near optimal solutions for problems by iteratively manipulating current solutions. After the old solution is manipulated, the new solution is compared to the old solution using an acceptance criterion to determine the quality of each solution. The solution with the higher quality, which is the greater value in the case of a maximization problem, is then used for the next iteration. When the solution values converge, the process ends.

One of the key merits of metaheuristics is that little or no assumptions are made about the problem. A sub-field of metaheuristics population-based optimization techniques which often draw inspiration form nature. According to Karaboga and Akay (2009), two important classes of population-based optimization algorithms are evolutionary algorithms and swarm intelligence-based algorithms. Popular evolutionary algorithms include Genetic Algorithms, Genetic Programming, Evolution Strategy and Evolutionary Programming.

As Wenping and Akay (2009) state, Swarm intelligence (SI) algorithms are innovative methods used to solve complex optimization problems. According to Bonabeau *et al.* (1999), SI algorithms are "any attempt to design algorithms or distributed problem-solving devices inspired by the collective behaviour of social insect colonies and other animal societies". Over the last two decades, the field of SI have received much attention and a number of algorithms were developed. Some algorithms include:

1. Ant Colony Optimization
2. Particle Swarm
3. Bacterial Foraging Optimization
4. Artificial Bee Colony

Karaboga and Akay (2009) state that the Artificial Bee Colony (ABC) algorithm performs better or similar to other SI algorithms. Carstens *et al.* (2011) developed a variation of the ABC algorithm that can be used for reliability modelling due to its good performance relative to other metaheuristic algorithms, its efficient multivariate problem solving ability and ease of use.

4.5.1 Artificial Bee Colony Optimization

The ABC was developed by Karaboga (2005) and is based on the foraging behaviour of honey bee colonies. The ABC model classifies bees into three groups: employed-, onlooker- and scout bees. Next, the basic logic of the ABC algorithm is discussed.

The algorithm starts of by finding 1000 stochastic solutions. A solution is a solution to Equation 4.4.9 obtained with the β, η and $\overline{\gamma}$ model parameters. Finding a stochastic solution entails generating a random value for each model parameter within a predefined search space. Each model parameter is given a search space within which the random value is found. Now, the size of the regression coefficient vector $\overline{\gamma}$ is determined by the amount of covariates included in the model.

If five covariates are used in the model, $\overline{\gamma}$ is a row vector containing five coefficients. Random model parameter values are generated using the equation given by:

$$p_{new_{rand}} = p_{min} + \phi(p_{max} - p_{min}) \tag{4.5.1}$$

where $p_{new_{rand}}$ indicates the new parameter value, p_{min} the minimum parameter value, ϕ a random value between [0,1] and p_{max} the maximum parameter value, proposed by Karaboga and Akay (2009).

After a 1000 stochastic solutions have been generated, the ten solutions with the highest quality are selected for further manipulation. The quality of a solution is a measure of the value of a solution. In the case of a maximization problem, the greater value of two solutions has higher quality. Generating 1000 solutions ensures that a thorough search of the algorithm search space is done which exhibits the explorative nature of the algorithm. In the case of a maximization problem, a good explorative algorithm "tests" its' entire search space to ensure that the solutions obtained aren't local maximum values.

Other than that, generating a large number of solutions across the entire search space will in most cases result in good initial model parameters and consequently high quality solutions. Next, three consecutive search processes are initialized. Each of these processes are associated with a bee as shown next.

The process of the employed bees start where each bee takes one of the ten stochastic solutions memorizing the solution and corresponding model parameter values. Each model parameter is then manipulated using the following equation:

$$p_{new_{search}} = (p - r_b) + (2 \cdot r_b)(\phi) \tag{4.5.2}$$

where $p_{new_{search}}$ represents the manipulated model parameter value, p the current model parameter value, r_b the search radius of the particular bee and ϕ a random value between [-1,1]. This process was proposed by Karaboga and Akay (2009) calling it the *bee dance* due to the fact that the bee "dances" around the solution in search of a higher quality solution. Each search process uses the *bee dance* to obtain modified values for each of the model parameters.

These model parameters are then fed into Equation 4.4.9 whereby a modified solution is obtained. If this modified solution value is of higher quality than the current solution value, the modified solution value and corresponding model parameter values are memorized, replacing the current values. If however the current solution value has higher quality than the modified solution, the modified solution and corresponding model parameter values are deleted.

To summarize, each of the ten employed bees takes a solution, modifies it memorizing the higher quality solution and its' corresponding model parameters. These ten solutions are then sent to the onlooker bees. Each bee performs the *bee dance* as the employed bee, but with a reduced search radius. Reducing the search radius is done to ensure that the local maximum is exploited. As with the employed bee, each of the ten onlooker bees compare their modified solution value with the current solution value memorizing the higher quality solution value and its' corresponding model parameter values.

An modification of the original ABC was made by Carstens *et al.* (2011) whereby a fourth bee, called the improvement bee, was added. The search process of the improvement bee is the same as with first two bees which does the *bee dance* with each of the ten given solutions memorising the higher quality solution and its' corresponding model parameter values. Initially, the search radius of the improvement bee is the same as that of the employed bee. A more detailed discussion of the improvement bee is done later.

Next, the scout bees generates ten random new solutions using Equation 4.5.1. This ensures further exploration of the search space during each iteration additionally adding to the algorithm's exploration capability. These solutions are then compared to the ten solutions found by the improvement bees, memorising the higher quality solution value and its' corresponding model parameter values.

This ends the search processes of an algorithm iteration, resulting in ten solution values and its' corresponding model parameters values. For the following iteration, these ten values are sent back to the employed bees at the start of the algorithm.

Coming back to the discussion on the search radius of the improvement bee, each search process rely on the search radius of the bee to find a higher quality solution. A small search radius might exploit the local maximum without being able to find the global maximum. If however a large search radius is used, the local maximum might not be found. To address this problem, Carstens *et al.* (2011) introduced the improvement bee which has a reducing search radius.

A reducing search radius ensures that the algorithm's explorative and exploitative ability is enhanced to obtain improved near optimal solutions. A test is used to decide whether the search radius of the improvement bee should be reduced. This is, whenever 5 consecutive iterations have equal solutions, its' search radius is reduced.

Lastly, two tests are done to determine whether the solutions obtained have converged. The first test stops the algorithm if the improvement bee search radius is less than 10^{-20}. If this condition is met, it is believed that the search radius is too small to make any substantial improvements to the solution. Finding a higher quality solution this way is time consuming as it is almost completely explorative and weak in terms of exploiting the local maximum.

As for the second test, the number of iterations are evaluated against the iteration limit of 5000. This ensures acceptable amount of time is spent to find solutions. These two tests are performed after the search processes of the four bees in each iteration. If any one of the conditions are met, the algorithm is stopped.

4.5.2 Evaluation of Artificial Bee Colony Optimization Success

Before the algorithm could be used to estimate the model parameters, the performance of it had to be evaluated. The given problem is that Equation 4.4.9 has to be maximized to estimate the model parameters. A case study was done whereby the algorithm was applied to determine model parameters for given data.

Figure 4.2 illustrates an algorithm run and how the solutions converged to a solution of -60.1001 after 1843 iterations. Data typically found in industry was used to develop the algorithm.

On average, it took the algorithm 1830 iterations before a stopping criteria was met. To test the variance in the solutions obtained with the algorithm, it was run for 50 times. It was assumed that the 50 solutions are a sample set taken out of an infinitely large population of solutions. The first test was done to determine the 95% Confidence Interval (CI) of the sample set using:

$$\left(\overline{x} - t_{n-1,0.025}\frac{s}{\sqrt{n}}; \overline{x} + t_{n-1,0.025}\frac{s}{\sqrt{n}}\right) \tag{4.5.3}$$

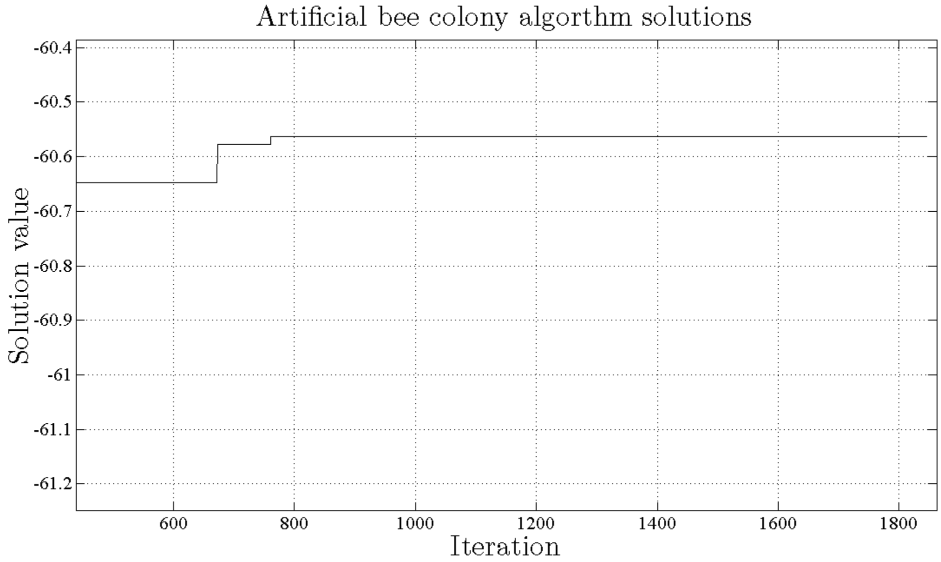

Figure 4.2: ABC algorithm solution plot.

where $\overline{x}$ represents the sample mean, n sample size, $t_{n,0.025}$ is found from a t-distribution table and s is the standard deviation as shown by Heiman (2010). This gave a sample mean of -60.2419 and $t_{49,0.025} = 2.0102$. As a result, the 95% CI was calculated as:

$$(-60.2764; -60.2074)$$

This indicated a fairly small variance in the sample solution values. However, although the variance seemed small it was necessary to determine the significance thereof. In order to do this, a 1000 stochastic solutions were generated. These 1000 values had a standard deviation of 45.6942×10^{22} whereas the sample set solutions had a standard deviation of 0.112. The variance in the 1000 solutions was then compared to the variance in the sample set to indicate the contrast using the F-test.

The F-test was found to be the most appropriate test to perform to compare the two sets of solutions. The F-test is based two hypotheses:

- H_0: There is no difference between the variances.
- H_A: The variances differ significantly.

where H_0 is rejected if the obtained F value is greater than the critical value. With a 95% confidence level, the corresponding critical value is 0.05. The F-test was applied to the two data-sets resulting in a F value of 1.63×10^{49} far greater than the critical value. As a result, there is not enough evidence to indicate that the two data sets do not vary significantly which meant that $H_(0)$ was rejected.

It can therefore be concluded that the variance in the sample set solutions are insignificant and solutions obtained with the algorithm are assumed to be near optimal. Subsequently, the model parameters corresponding to the sample solutions were also near optimal which meant that the algorithm can be used for this thesis.

4.6 Goodness-of-fit tests

Goodness-of-fit (GOF) tests are used to determine how good a model fits a certain data set. More specifically, these tests measure how a data set compares to a theoretical probability density function. Two types of GOF test are graphical- and analytical-tests. These tests, apart from testing the model fit, evaluates the PHM assumptions upon which the model is based.

The first PHM assumption, given in Chapter 3, can be evaluated using graphical techniques (Pham, 2006). The basic principle of graphical tests, according to Miller *et al.* (1981), is that the scales of the axes should be chosen such that if the model holds, a straight line plot represents the data. On the other hand, if the model fails, the plot resembles a curve.

Evaluating the second PHM assumption can be done using graphical- and/or analytical-tests. The general procedure for analytical tests entail specifying a test statistic. A test statistic is a function of the data set which represents the distance between the theoretical probability density function, sometimes referred to as the hypothesis, and the data set.

To test the model adequacy, the probability of obtaining a data set with a greater test statistic than the current has to be determined. A greater test statistic would mean that the model does not fit the data sufficiently. Probabilities close to unity are rarely found and can be an indication that mistakes were made during testing procedures.

Table 4.1: Goodness-if-fit Tests

Analytical	**Graphical**
Kolmogorov-Smirnof	Cox-Snell
Anderson Darling	Martingale

Several GOF tests can be used for these evaluations. For the estimation of Weibull PHM model parameters, the discussion of GOF tests is limited to the tests tabulated in Table 4.1.

4.6.1 Analytical Techniques

The basic principle of analytical tests are defined by the following two hypothesis:

- H_0: The data follows a specified distribution $F_n(x) = F_0(x)$.
- H_1: The data does not follow the specified distribution $F_n(x) \neq F_0(x)$.

where $F_0(x)$ is the probability density function (PDF) and $F_n(x)$ the empirical distribution function (EDF), or the hypothesis function. The tests discussed in this section are: Kolmogorov-Smirnoff test and the Anderson-Darling test.

4.6.1.1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) test, sometimes referred to as the *d*-test, is a non-parametric test and assumes a continuous distribution. Also, Milbrodt and Strasser (1990) states that it is less sensitive at the tails of the distribution than at the center of the distribution. A procedure for the test was set out by Bekker (2011) and is repeated here using several steps.

Let $X_i = [X_1, X_2, X_3, ..., X_n]$ denote the ordered survival times of the events. The test is introduced stepwise. Consider the following:

1. Null hypothesis H_0: The data has a theoretical distribution $f_0(x)$.
2. The empirical distribution function is given by:

$$F_n(x) = \frac{number\ of\ \ X_i's \leq x}{n} \quad \texttt{for} \quad i = 1, 2, ..., n. \tag{4.6.1}$$

3. The K-S test statistic, which determines the largest vertical distance between the two functions at each X_i, is given by:

$$D_n = max|F(x_i) - F_0(x_i)| \tag{4.6.2}$$

   and D_n is calculated using:

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F_0(x) \right\} \tag{4.6.3}$$

$$D_n^- = \max_{1 \leq i \leq n} \left\{ F_0(x) - \frac{i-1}{n} \right\} \tag{4.6.4}$$

$$D_n = \max \left\{ D_n^+, D_n^- \right\}. \tag{4.6.5}$$

4. If D_n exceeds some critical value, the null hypothesis is rejected. This means that when the vertical distance between the two distributions exceeds a certain distance, the null hypothesis is rejected. The test condition is given by:

$$D_n > c_{nj\alpha} \tag{4.6.6}$$

The critical value, $c_{nj\alpha}$, is obtained from the K-S One-Sample statistic table.

4.6.1.2 Anderson-Darling Test

The Anderson-Darling (A-D) test is a modification of the K-S test and was introduced by Andersen (1982). Compared to the K-S test, it has the advantage of placing more emphasis on the tails of the distribution. Additionally, the critical values do not depend on the distribution being investigated.

However, specific distributions each have critical values calculated for it. As a consequence, the A-D test is a more sensitive test compared to the K-S test. The key limitation of the model is the computational effort necessary to calculate critical values for each specific distribution. For this test the distance between the PDF and the EDF is given by:

$$n \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 w(x) dF(x) \tag{4.6.7}$$

where $w(x)$ is a weighting function. According to Anderson and Darling (1954), if the weighting function is given as:

$$w(x) = [F_0(x)(1 - F_0)]^{-1} \tag{4.6.8}$$

the distance between the PDF and EDF is given by:

$$n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{[F_0(x)(1 - F_0(x))]} dF(x) \tag{4.6.9}$$

In order to test the model fit, a test statistic has to be determined which is defined by:

$$A^2 = -n - S \tag{4.6.10}$$

where

$$S = \sum_{i=1}^{n} \frac{2i-1}{n}[\mathtt{ln}F(X_i) + \mathtt{ln}(1 - F(X_{n+1-i}))] \tag{4.6.11}$$

Due to the computational effort of calculating critical values for each distribution, tabulated values have been published for specific distributions. As with the K-S test, the A-D test is also an one-sided test and the hypothesis is rejected if the test statistic, A, is larger than the critical values.

4.6.2 Graphical Techniques

Several different graphical techniques are found in literature to assess the adequacy of a model fit. Two of these techniques are cumulative hazard plots and residual hazard plots. The tests discussed in this section are: Cox-Snell residuals and Martingale tests. Both of these tests are combinations of both cumulative hazards plots and residual plots.

4.6.2.1 Cox-Snell Tests

The original test was introduced by Cox and Snell (1968) and is also known as the Cox-Snell residual test. Consider the cumulative FOM given by:

$$H(x) = -\log(R(x)) = -\log(1 - F(x)) \tag{4.6.12}$$

where $R(x)$ is the reliability function, $F(x)$ the cumulative failure distribution function and $H(x)$ the cumulative FOM. Let $\overline{z(x)}$ denote the i^{th} item's covariate vector. According to Klein and Moeschberger (2003), the true cumulative FOM for an item failure is denoted by $H(X_i|\bar{z}(x))$.

Once the PHM has been fit to the data and the values for $\overline{\gamma}$ have been estimated i.e., $\overline{\gamma} = (\gamma_1, \gamma_2, \gamma_3, ..., \gamma_n)^t$, the cumulative FOM can be given as:

$$H(X_i, \overline{z_{im}(x)}) = H_0(X_i)exp\left[\sum_{i=1}^{p} \overline{z_{im}(x)} \cdot \bar{\gamma}\right] \tag{4.6.13}$$

where failures are denoted by $i = 1, 2, 3, ..., n$, the maintenance inspection interval denoted by p and the particular covariate by m. Further the Cox-Snell residuals are defined for time dependent covariates as follows:

$$r_{CSi} = h_0(X_i) \exp\left[\sum_{k=1}^{p} \overline{z_{im}(X_i)} \cdot \bar{\gamma}\right], i = 1, 2, ..., n. \tag{4.6.14}$$

Here, h_0 is the baseline FOM given in Equation 6.1.1. To check the validity of the model, Tableman and Kim (2004) state that the r_{CSi}'s should resemble a censored sample from a unit exponential distribution.

To explain this, suppose $H(x) = x$ has a cumulative FOM of the unit exponential distribution. The estimate of the cumulative FOM versus r_{CSi} should therefore exhibit a straight line through the origin with a slope of one. To illustrate a Cox-Snell plot, dummy data was generated which illustrates a good model fit in Figure 4.3.

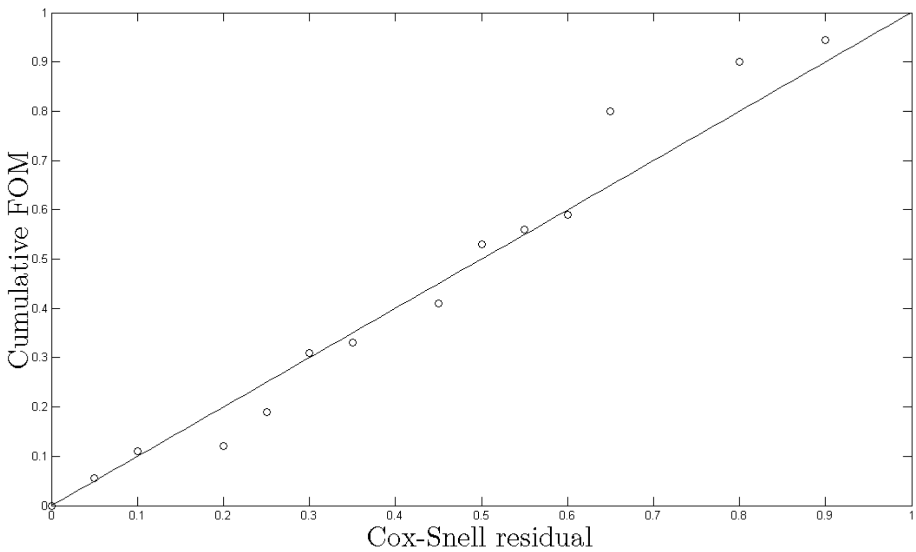

Figure 4.3: Typical Cox-Snell residual plot.

The scatter plot arround the line through the origin indicates that the model fits the data sufficiently. A limitation of the model is that when the cumulative hazard plot is not linear, the test cannot show the type of departure from the model. Also, departures from the unit exponential distribution might be due to the uncertainty in estimating H_0 and the regression coefficients $\bar{\gamma}$. These uncertainties are found with small samples and are largest in the right-hand tail of the distribution.

4.7 Decision Model

This section deals with the development of a decision model based on Residual Useful Life (RUL) and cost optimization. In the case of non-repairable systems, RUL and cost optimization estimations are based on the observed FOMs discussed in Section 4.4. These estimates are used to develop a decision model.

Several authors have developed decision such models based on cost optimization and the relationship between RUL and the FOM (conditional probability of failure), for example, Jardine (1973), Newby and Dagg, Pintelon and Gelders (1992), Wu and Ryan (2011), Lugtigheid *et al.* (2007), Makis and Jardine (1992),Newby (2008) and Banjevic *et al.* (2001).

Jardine (1973) presented decision models taking into account factors such as operating cost, long term cost minimization, technological improvements, items prone to breakdown and minimization of downtime. Conventional renewal methods attempt to minimize the long term operational cost of an item by finding an minimum cost level, known as the minimum long term Life Cycle Cost (LCC).

In order to determine the minimum long term LCC, two decision rules are discussed in this section, namely (a) decision making with RUL estimation and (b) decision making aided by cost optimization.

4.7.1 Residual Life Estimation

Calculation of the RUL is based on the FOMs discussed in Section 4.4.1, i.e. Equation 6.1.1. Authors mentioned in the introduction of this section utilize the conditional event expectation of an item to determine RUL estimates. For this reason this section deals with the conditional failure expectation.

This section makes use of three reliability functions, i.e. $f(x)$, $F(x)$ and $R(x)$. To include the influence of covariates, the functions have to functions of time and covariates, i.e. $f(x, \overline{z(x)})$, $F(x, \overline{z(x)})$ and $R(x, \overline{z(x)})$. For generality, assume time-dependent covariates.

Suppose a decision policy for an item at time x states that the item should be replaced at time X_p or at failure, whichever comes first. The conditional expectation, $X_{r+1} \leq X_p$, can then be given by:

$$E[X_{r+1} \,|X_{r+1} \leqslant X_p] = \frac{\int_x^{X_p} x \cdot f_X(x, \overline{z(x)})dx}{\int_x^{X_p} f_X(x, \overline{z(x)})dx} \tag{4.7.1}$$

where the expected RUL to the $(r+1)^{th}$ event can be given by:

$$\mu_{r+1} = E[X_{r+1} \,|X_{r+1} \leqslant X_p] - x \tag{4.7.2}$$

A confidence interval can be estimated around the RUL in (4.7.1) to represent the certainty of the estimate. Let the upper confidence limit of the expected residual life be denoted by $\tilde{X}_{r+1}$, and the lower limit by $\underset{\sim}{X}_{r+1}$. Calculation of the upper confidence level can be done by numerically solving:

$$\frac{\int_x^{\underset{\sim}{X}_{r+1}} f_X(x, \overline{z(x)})dx}{1 - \int_0^x f_X(x, \overline{z(x)})dx - \int_{X_p}^{\infty} f_X(x, \overline{z(x)})dx} = 1 - \frac{1 - CI}{2} \tag{4.7.3}$$

where CI indicates the confidence interval. The lower limit is calculated similarly by numerically solving:

$$\frac{\int_x^{\underset{\sim}{X}_{r+1}} f_X(x, \overline{z(x)})dx}{1 - \int_0^x f_X(x, \overline{z(x)})dx - \int_{X_p}^{\infty} f_X(x, \overline{z(x)})dx} = 1 - \frac{CI}{2} \tag{4.7.4}$$

However, in some cases it is required to determine the RUL when no preventive maintenance rule is present, i.e. $X_p = \infty$. In this case the expected residual life at $x \approx 0$ can be calculated using a modified version of Equation 4.7.1 given by:

$$E[X_{r+1}] = \frac{\int_0^{\infty} x \cdot f_X(x, \overline{z(x)})dx}{\int_0^{\infty} f_X(x, \overline{z(x)})dx} \tag{4.7.5}$$

and is known as the non-repairable system's Mean Time Between Failure (MTBF). The expected RUL to the $(r+1)^{th}$ event can be given by:

$$\mu_{r+1} = E[X_{r+1}] - x \tag{4.7.6}$$

with the upper confidence interval equation given as:

$$\int_0^{\tilde{X}_{r+1}} f_X(x, \overline{z(x)})dx = 1 - \frac{CI}{2} \tag{4.7.7}$$

and the lower confidence interval equation given as:

$$\int_0^{\underset{\sim}{X}_{r+1}} f_X(x, \overline{z(x)})dx = 1 - \left(1 - \frac{CI}{2}\right) \tag{4.7.8}$$

To visualize RUL, a graphical illustration of the concepts is given in Figure 4.4. Suppose that the CM inspection interval for a specific non-repairable item is X_c. At each interval the RUL and corresponding confidence interval is indicated by a

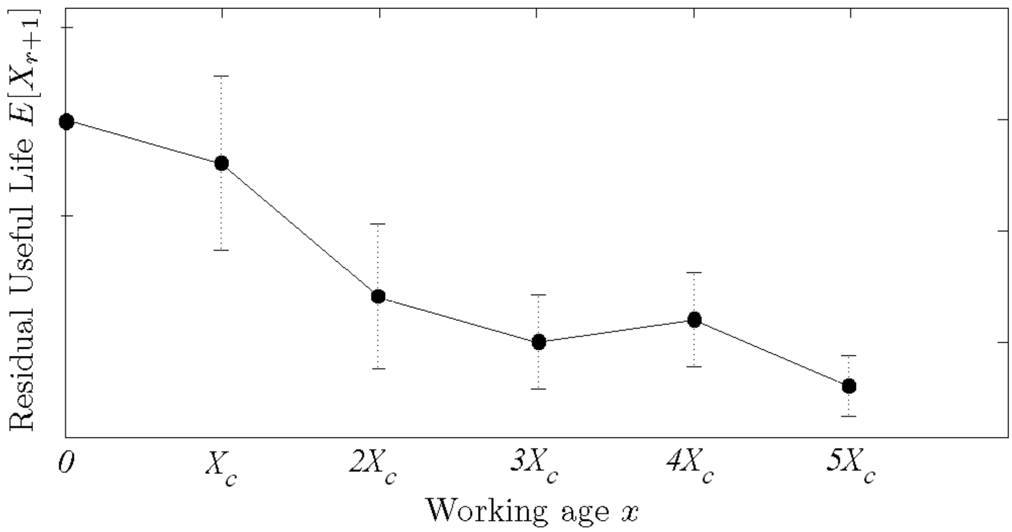

Figure 4.4: Illustration of RUL of non-repairable items.

dot and the dotted line, respectively.

In theory, the RUL decreases linearly by X_c days in an interval of X_c days. Although this might be valid in theory, in real life scenarios minor maintenance actions are performed, the life of an item might be extended, or decreased. When these maintenance actions are performed, the covariate influence is changed.

For example, suppose a pump is misaligned by 0.5 degrees with axial vibration measurement of $0.76mm/s$ with a certain RUL. If maintenance is performed X_c days later and the misalignment is corrected, the vibration measurement might reduce to $0.2mm/s$. As a result, the RUL is increased. In figure Figure 4.4 this scenario is illustrated after the inspection at $3X_c$.

It is evident that the covariates have a direct affect on the RUL of an item. In most cases data obtained from industry contains many errors and gaps in the data which has to be filled. For this reason, assumptions about the covariates have to be made. These assumptions are discussed in Chapter 5.

4.7.2 Decision Making Rule With Residual Life Estimation

It might seem obvious to take maintenance action when the lower limit of the RUL estimate is zero. A few problems are associated with this. Firstly, inspection intervals are performed discretely. However, if the lower limit was close to zero at the inspections interval, it might decrease to zero between maintenance intervals placing the item at high risk of failure.

Secondly, RUL estimations are derived from the conditional failure expectation which has the consequence that it only reaches zero as time approaches infinity. To overcome these problems, a decision rule is given as follows: maintenance action is be taken as soon as the lower RUL limit minus the inspection period divided by two is less or equal to the time when the estimate is determined. This is represented as follows:

$$\underset{\sim}{X}_{r+1} - X_c/2 \leq x \tag{4.7.9}$$

where X_c the interval.

4.7.3 Cost Optimization Models

The objective of cost optimization models are to minimize the long term LCC of an item. Two factors are balanced: (a) cost of preventive maintenance and (b) cost of unexpected failure. Preventive maintenance is when maintenance is performed in attempt to prevent the occurrence of failure. Renewal maintenance is when an item has already failed and need to be replaced. In Figure 4.5 these factors are plotted indicating the optimal instant when to perform maintenance.

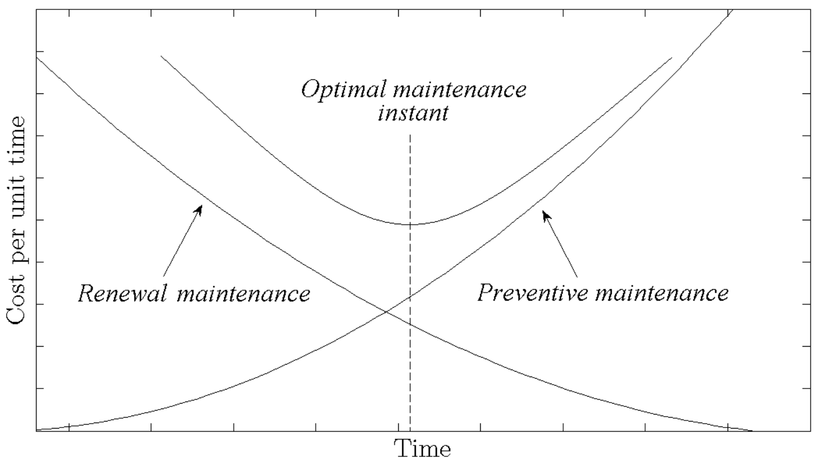

Figure 4.5: Optimal maintenance instant.

Let C_f indicate the cost of unexpected failure replacement and C_p the cost of preventive replacement. Also, the time required for preventive replacement is represented by a, and the time required for failure replacement is represented by b. The total cycle cost of maintenance can be given by:

$$C_{Xp} = C_p \cdot R_X(Xp, \overline{z(x)}) + C_f \cdot F_X(Xp, \overline{z(x)}) \tag{4.7.10}$$

where R_X and F_X are defined in Chapter 2 as the reliability function and cumulative distribution function, respectively. It should be noted, even when an optimum maintenance instant X_p is defined, expected failures might still occur. Furthermore, the expected duration of a non-repairable item's life cycle at time $x = 0$ is given by:

$$E[X_p] = (X_p+a)\cdot R_X(X_p, \overline{z(x)}) + (E[X_{r+1}|\, X_{r+1} \leqslant X_p]+b)\cdot F_X(X_p, \overline{z(x)}) \tag{4.7.11}$$

Dividing Equation 4.7.10 by Equation 4.7.11 yields the cost of maintenance per unit time:

$$C(X_p) = \frac{C_p \cdot R_X(X_p, \overline{z(x)}) + C_f \cdot F_X(X_p, \overline{z(x)})}{(X_p + p) \cdot R_X(X_p, \overline{z(x)}) + (E[X_{r+1} \mid X_{r+1} \leq X_p] + f) \cdot F(X_p, \overline{z(x)})} \tag{4.7.12}$$

where the minimum cost is found where $dC(X_p)/dx = 0$. Detailed inference of the model is given by Jardine (1973). Dummy-data was generated to plot $C(X_p)$ in Figure 4.6 which indicates the minimum cost per unit time located near.

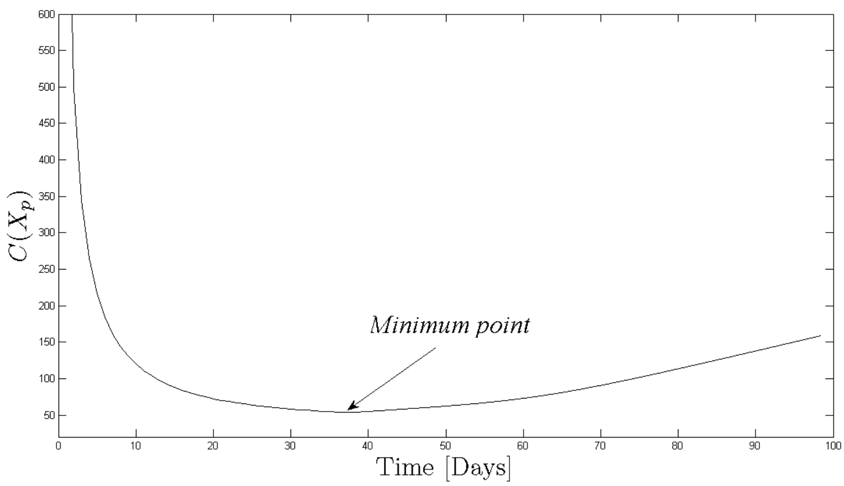

Figure 4.6: Long term cost per unit time optimization.

4.7.4 Final Decision Model

At this moment, two estimations have been introduced, namely: (a) Residual Useful Life and (b) long term cost optimization. It is not known which of the two estimations would produce the best estimations for the event arrival time. These decision models are presented here with an additional decision model.

- Decision model 1: Perform maintenance according to the residual useful life estimates, i.e.

$$\underset{\sim}{X}_{r+1} - X_c/2 \leq x \tag{4.7.13}$$

  where $\underset{\sim}{X}_{r+1}$ is the lower confidence limit, X_c the inspection interval and x the current oil age in hours.

- Decision model 2: Perform maintenance according to the optimal point on the cost function.

$$x = X_p \tag{4.7.14}$$

  where x is the current point in time and X_p the recommended preventive replacement time.

- Decision model 3: Combine decision model 1 and 2 and perform maintenance whenever either of the two decision models suggests that maintenance has to be performed.

The decision models are applied to the case study in Chapter 5. The results are then compared and decided which model produces the best event arrival time predictions.

4.8 Conclusion

In Chapter 4 two different forms of the PHM were presented, e.g. semi-parametric PHM and fully-parametric PHM. The fully-parametric Weibull PHM was found to be the most suitable form identified for this thesis. An algorithm developed by Carstens *et al.* (2011) was found to be acceptable for parameter estimation of the fully-parametric Weibull PHM. Several Goodness-of-fit models were introduced with which the model parameters could be tested. Lastly, two decision models were introduced whereby a third model was given which is a combinations of these two decision models.

Chapter 5

Proportional Hazards Model Application

5.1 Introduction

The Proportional Regression Model (PHM) investigated in Chapter 4 is applied to data obtained in the South African mining industry which is then also used to validate the theory discussed in Chapter 4.

Most industries in South Africa do not know how reliability modelling can improve overall asset performance. As a result, there is a lack of commitment towards the recording of good quality data which cannot be used for reliability modelling.

Because data quality is such a problem, it was necessary to develop data requirements which can be used to evaluate data to determine whether it is suitable for reliability modelling. After a long search of suitable data, data found at Sishen satisfied these requirements. Sishen is an open pit iron ore mine in the Northern Cape and is part of Anglo American (Kumba Iron Ore).

Sishen has several assets which have sophisticated Condition Monitoring (CM) equipment implemented on it. One of the crucial assets to Sishen's operation is their fleet of 78 haul trucks. As a result, it was decided to apply this thesis to data obtained from these haul trucks to illustrate the benefit of reliability modelling.

Currently there are three truck manufacturer's models in operation, namely: Caterpillar , Komatsu and Liebherr. The Caterpillar (CAT) trucks employ mechanical drive powertrains. Komatsu and Liebherr on the other hand employ diesel-electric powetrains with a diesel generator which powers electric wheel-motors at each wheel.

It was found that the CAT trucks had the most readily available data and was therefore chosen for this study. Sishen has eight CAT 793D trucks in operation. These are 78l diesel engines capable of producing 1742kW and moving 218 ton payload. Each truck has an on-board Computerized Maintenance Management System (CMMS) which monitors and records several conditions of the truck. Also, Sishen has a state-of-the-art tribology center which analyzes oil samples taken weekly from each truck.

This data was evaluated using the data requirements and is discussed Section 5.2. Once it was established that the data was suitable, analysis and modelling thereof was done using the theory discussed in Chapter 3. This is discussed in Section 5.4. The decision model is then applied and discussed in Section 5.5.1.

5.2 Data Requirements

The requirements given in this section was developed to evaluate data to determine whether it is suitable to be used for Proportional Hazards Modelling modelling. For the sake of the discussion, it is assumed that the process parameter is time, such as hours. Other process parameters are discussed in Chapter 2.

5.2.1 Events

Firstly, it is necessary to define an event. Events might include any of the following:

- Preventative and scheduled maintenance: Typically, this is minor maintenance actions such as lubrication and realignment which affects the influence that the covariates has on the FOM.
- Failure replacement: When an item or system is replaced.
- Censored items: This might be an item which has been taken out of operation or not have not yet failed at the time when the observation period ended.

Events are when maintenance action is taken which influenced the item survival time. Given this information, it is important to know the time when any of these events occurred. It is also necessary to know when the item was put into operation. Lastly, to develop a meaningful PHM, a minimum of approximately 15 events are required. In order for covariates to converge, sufficient events are necessary with which the PHM is constructed.

5.2.2 Covariates

Covariates are explanatory variables that may be predictive of the outcome under study. Two types of covariates were introduced in Chapter 2, namely: (a) *quantative*-covariates and (b) *subjective*-covariates.

Quantative covariates Condition Monitoring (CM) recordings which include (but are not limited to): temperature, vibration, pressure, oil content and stress. Quantitative covariates may be time-dependent or time-independent and can either be measured continuously or measured at a fixed intervals. These covariates should be recorded at each possible instant and a recoding is necessary at each event occurrence.

Subjective may covariates are binary variables representing factors such as: whether oil changed at each service interval, type of oil used, maintenance team used, type of installation set up, supplier used and operator on shift. Depending on the data, this is basically comprises of yes/no data.

One method to obtain subjective data can be to speak with the technical personnel that maintain and run the items. With their knowledge and experience it is easier find the correct subjective covariates which contributes to the occurrence of events. Another method to obtain subjective data through CMMS systems with which the data can be recorded.

5.2.3 Requirements

In conclusion, the main requirements of data to perform Proportional Hazards Modelling is:

- Time: The time when the item was put into operation and the time when events occurred.
- Covariates: Recordings of all possible covariate values should be done at the time of each event. It is also preferred that covariate values are recorded between events. Covariate recordings made between event occurrences, may be done continuously and/or discretely, depending on the CM equipment installed and maintenance strategy followed.

These requirements are necessary to develop a PHM. Later in this chapter it is shown how assumptions about data can bare made to adhere to these requirements. It should therefore be noted that if given data does not exactly fit these requirements it does not necessarily mean that it cannot be used. For this reason these data requirements should merely be used as guidelines for data to develop a PHM.

5.3 CAT 793D Data

Data was obtained from the CAT 793D haul trucks operating at Sishen. Figure 5.1 illustrates an CAT 793D haul truck fully loaded. Two types of diagnostic of data recorded namely: a.) on-board Condition Monitoring measurements and b.) results of oil analysis sample data.

Figure 5.1: Cat 793D haul truck.

For the purpose of this thesis, all the conditions related to the diesel engine condition was used for PHM development. Therefore, from this point forward data related to the condition of the related to the engine of the haul trucks are taken into account and all other data is ignored.

5.3.1 On-Board Measurements

Each truck has an Electronic Control Module (ECM) which utilizes engine management software to monitor, protect and control the engine utilizing its' diagnostic sensors to measure several important conditions of the truck. These sensors take into account operating conditions and power requirements to adapt the engine such that maximum performance and efficiency are achieved. On board measurements used are:

- Air filter pressure (max.) [kPa].
- Ambient air temperature (avg.) [°C].
- Boost pressure (avg.) [kPa].
- Engine coolant temperature (avg.) [°C].
- Oil filter pressure (max.) [kPa].
- Oil filter pressure (min.) [kPa].
- Left exhaust temperature (max.) [°C].
- Right exhaust temperature (max.) [°C].
- Right after cooler temperature (max.) [°C].

On-board measurements are made hourly whereby the maximum or average for that hour is taken. The ECM has "alarm" triggers which sound when a certain condition reaches an unacceptable level. Additional to the alarm signal in the cockpit, a signal is sent to a remote computer at the maintenance center. These alarms can be bypassed to shut them off, ignoring the alarm. Switching off these alarms might be attributed to one of two reasons.

Firstly, the driver is busy loading the truck and cannot stop the shift to take the truck to maintenance. Secondly, the driver switches the alarm off because they do not believe it would lead to failure.

If the alarms are ignored it places the haul truck at risk of engine failure. If the engine fails, the haul truck is taken out of service for a period of time far greater than that of the time required for preventive replacement. Due to the dependency upon the haul trucks for operation, taking a haul truck out of service for an extended period of time has major financial implications. For this reason, the necessary measuremens should be taken to avoid the occurrence of failure.

With this being said, a haul truck consists of several thousand components. For this reason the ECM cannot precisely indicate which component is faulty thereby placing several components at risk of failure if an alarm is ignored. This further motivates the need to develop a model which can be used to estimate when failure is likely to occur. This in turn would reduce the maintenance engineers' uncertainty about the failure behaviour of the haul trucks.

5.3.2 Oil Analysis

Oil samples are taken weekly whereby analysis thereof provides the constituents in the oil in parts per million (ppm). Some of the constituents found in the oil are:

- Iron [Fe]
- Chrome [Cr]
- Lead [Pb]
- Copper [Cu]
- Silica [Sn]
- Aluminium [Al]
- Nickel [Ni]
- Gold [Ag]
- Silicone [Si]
- Barium [Ba]
- Nitrous [Na]
- Magnesium [Mg]
- Calcium [Ca]
- Zinc [Zn]

Different forms of viscosity are also determined when oil analysis is done. These include:

- Viscosity I [VI]
- Viscosity 40 [V40]
- Viscosity 100 [V100]

According to the engineers at the tribolgy center, some of the most influential elements found in the oil are: Fe, Si and V100. These elements can be used to determine, to some extent, how the engine is deteriorating. For example, degradation of components such as cylinders, gears, pistons, shafts and valves is indicated by the Fe found in oil.

Furthermore, due to the dusty environment the haul trucks operate in, dust is sucked up through the air filters. The dust contains a large amount of Si and high levels present in the oil might indicate that the air filters or gasket seals are degrading and need replacement.

Lastly, V100 is an indication of the viscosity of the oil and can indicate fuel dilution, mixture of oils and water in oil. Fuel dilution is indicated by low V100 levels which means traces of diesel are found in the oil. When diesel is found in the oil a fuel component might be faulty and replacement thereof has to be done before failure occurs and other components are damaged.

These are three of the most important oil analysis measurements found to indicate degradation of the haul truck engines. Each measured value is compared to three predefined values, namely: maximum value, minimum value and maximum gradient. Values outside the predefined maximum and minimum values initialize maintenance inspection and/or oil change. If the gradient level is reached, the trend in the element value surpasses the predefined value indicating that a component is degrading faster than it should and maintenance teams need to determine the cause.

Oil samples not only contain the constituent values, it also contain the oil sample date and the age of the oil in hours. A major problem the tribology center is experiencing is that maintenance technicians are topping up the oil of the haul truck's when it is being fuelled. Adding new oil to old oil dilutes the old oil which has the consequence that oil analysis results are inaccurate. It wouldn't be a problem if the precise amount of oil and time of the oil addition was recorded but no such attempt is being made.

As a rule, the oil of a truck is changed after 750 hours. Therefore the maintenance tactic followed is a combination of Condition Based Maintenance and Usage Based Maintenance discussed in Chapter 2.

5.3.3 Maintenance Decisions

At Sishen, maintenance is mainly based on oil analysis results. On-board recordings are rarely used to determine maintenance instances. Maintenance is initialized if one of the three constituent limits are reached. Although these procedures might be set in place, it is a conservative method in that the constituents are taken into account individually. The problem with this is that failure might be brought upon by more than one constituent. For this reason, a prognostic model such as the

PHM is used to determine the correlation between several constituents and failure.

For this reason it is possible to include on-board recorded values in conjunction with oil analysis values to develop a PHM model which describes the underlying failure behaviour of the haul trucks.

5.4 Weibull PHM Fit

In order to determine the most appropriate model fit is an iterative process. In order to develop the best model fit, Goodness-of-Fit (GOF) test and technical experience gained during this project were used. The main objective of model development is to determine the most appropriate covariates for a good model fit.

The process of model development is given below.

- Data manipulation.
- Covariate selection.
- Goodness-of-fit tests.
- Final Model.

A detail discussion of these items are presented in the following sections whereby it is shown how the final Weibull PHM was developed.

5.4.1 Data Manipulation

In this section all the assumptions made about the data are discussed. The period for which data was obtained for each of the eight trucks was approximately from October 2010 to June 2011.

Once the on-board truck data and oil analysis were obtained, the next step was to combine the data. However, because on-board measurements are made hourly and oil samples taken weekly, it was necessary to "fill the gaps" between the oil sample data points to have oil data for each on-board measurement.

When the data was combined two problems were identified. Firstly, data between weekly oil samples had to be interpolated. To overcome this problem, the engineers at the tribolgy center suggested that between weekly oil samples should be interpolated linearly.

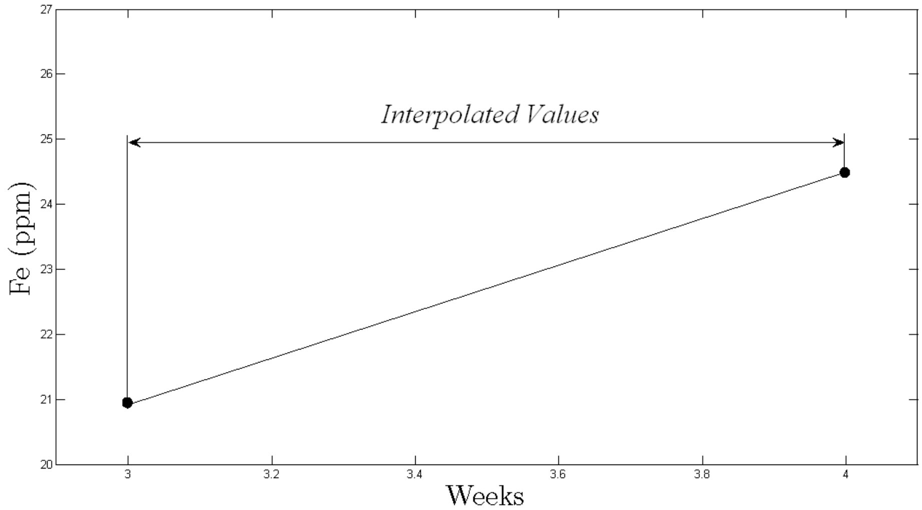

Figure 5.2: Extrapolation of oil data.

For example, suppose that oil samples were taken in week 3 and week 4 indicating Fe values of = 21 ppm and = 24.2 ppm. The data is then interpolated from week 3 to week 4 as shown in Figure 5.2. This meant that there was corresponding oil sample data for each on-board data measurement.

The second problem found was that oil changes had taken place between oil samples and it was not indicated when it took place. To solve this problem consider Figure 5.3. Suppose that the oil data is known at weeks 7, 8, 9 and 10. The oil age at these weeks are: 326, 446, 98 and 198.

Due to the fact that the oil age in week 10 is less than the oil hours in week 8, an oil change occurred between week 8 and 10. Data was then extrapolated linearly from week 10 backwards to the point where Fe is zero. This point was therefore where an oil change had taken place. Then, the Fe (ppm) was kept constant from week 8 to where the oil change took place. It was kept constant due to the uncertainty of which might have happened right before the oil change took place. Although this assumption might be conservative, the decision was based upon the knowledge of the engineers of the tribology center.

In order to validate the final model, 7 of the 8 trucks were used to develop the final Weibull PHM model. The model was then applied to the 8_{th} truck's data to validate its ability to model the failure behaviour of this haul truck.

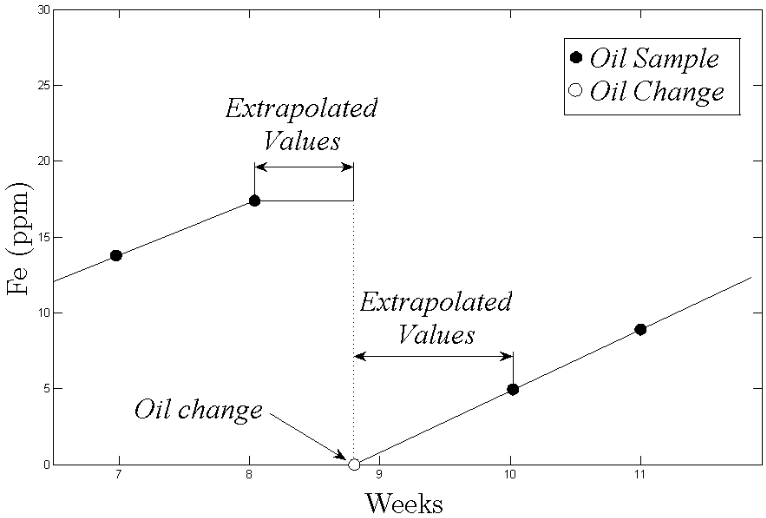

Figure 5.3: Extrapolation of oil data.

The haul trucks have similar failure patterns and deciding which haul truck to use for validation was trivial. It was decided to choose haul truck number 4 due the fact that the period over which data was available for this haul tuck contained sufficient observations to validate the model. Further discussions about the model validation is done in Section 5.5.

As for the other 7 haul trucks, 19 events were observed for the given period of time. Table 1 is a tabulation of these 19 events with the corresponding haul truck, failure time (X_i) and failure type (C_i).

In Table 1 it can be seen that 19 events were observed of which 4 were failures and 15 were censored observations. An event table which contains the 15 covariates at each event can be seen in Table A.1. Due to the vast amount of data used in this study it is not possible to include all of it in this thesis.

Table 5.2 is an extract of 4 of the 16^{th} event's measured covariates. Table A.2 is an extract of all of the 1^{st} event's measured covariates. Deciding which covariates to use for the model played an important part in this project and is discussed next.

5.4.2 Covariate Selection

Selecting covariates for this study was initially done based on the experience of the engineers at the tribolgy center and technicians at the maintenance workshop.

Table 5.1: Event table.

	Haul truck	X_i	C_i
1	1	552.00	0
2	2	418.32	0
3	2	561.00	0
4	3	634.67	0
5	3	538.00	0
7	4	530.00	0
8	4	384.00	0
9	4	444.00	0
10	4	341.00	0
6	4	445.00	0
11	5	486.42	0
12	5	385.00	0
13	5	872.00	0
14	6	454.00	0
15	6	399.00	0
16	6	904.00	1
17	7	562.00	0
18	7	222.18	1
19	7	422.64	0
20	7	338.00	0
21	8	840.00	0
22	8	48.40	1
23	8	105.60	1
24	8	324.00	0

The engineers are of the opinion that the most influential oil data covariates are: Fe, Si, Na, V100, Pb and Cu.

A combined total of 15 covariates were therefore used. All of these covariates were believed to contribute to failure in some way. To determine the most appropriate covariates for a good Weibull PHM fit, different covariate combinations have to be analyzed. In order to determine the quality of an model fit, Goodness-of-Fit (GOF) tests are used.

Table 5.2: An extract of the 16^{th} event's measured covariates.

#	x	C_i	Right After cooler temperature [°C]	Engine oil pressure [kPa]	Fe [ppm]	Si [ppm]
16	811.05	0	60.00	455.50	0.1934	3.7807
16	812.05	0	60.00	454.00	0.1936	3.7894
16	813.05	0	60.00	459.50	0.1939	3.7982
16	814.05	0	60.00	454.00	0.1942	3.8070
16	815.05	0	55.00	468.5	0.1944	3.8157

5.4.3 Goodness-of-fit Tests

Three GOF tests were discussed in Chapter 3, namely: Kolmogorov-Smirnoff, Anderson-Darling and Cos-Snell. Compared to the other two tests, the Kolmogorov-Smirnoff (K-S) test is more flexible and less complex to use. For these reasons the K-S test was the chosen GOF test for model fit evaluation.

The process of the K-S test was presented in Chapter 3 and is briefly repeated here for convenience. The basic principle of the K-S test is defined by the following two hypothesis:

- H_0: The data follows a specified distribution $F_n(x) = F_0(x)$.
- H_1: The data does not follow the specified distribution $F_n(x) \neq F_0(x)$.

The K-S test attempts to find the maximum vertical distance between the cumulative Probability Density Function (PDF) and the Empirical Distribution Function (EDF). This vertical distance is known as the test statistic, D_n, and is given by:

$$D_n = max|F(x_i) - F_0(x_i)| \tag{5.4.1}$$

The test statistic is compared to a critical value, $c_{nj\alpha}$, using the test condition given by:

$$D_n > c_{nj\alpha} \tag{5.4.2}$$

where $c_{nj\alpha}$ is obtained from the K-S One-Sample statistic table. If D_n exceeds the $c_{nj\alpha}$ there is not sufficient evidence to reject the null-hypothesis. With 19 events $n = 19$ and a confidence interval of 95% $\alpha = 0.05$ and $c_{nj\alpha} = 0.301$. Several different covariate combinations were evaluated. However, only the 3 most important combinations of interest is discussed in detail.

For each combination, the model fitting procedure discussed in Chapter 3 was applied to determine the model parameters, i.e. β, η and $\overline{\gamma}$. In each case the cumulative PDF is determined using these estimations. The EDF is determined using Equation 3.6.1 in Chapter 3.

Several covariate combinations were analyzed to determine the best combination to develop a Weibull PHM model which best describes the underlying failure behaviour of the item under study. The combinations which produced the best model fit is presented with its' corresponding model parameters and solution to Equation 3.4.9 in Chapter 3 maximized for parameter estimation.

5.4.3.1 Covariate Combination 1

The first combination consisted of Fe, boost pressure, left exhaust temperature and engine oil filter pressure. The calculated test statistic, D_n, is equal to 0.2129. According to Equation 5.4.2, the K-S test statistic for this combination is less than the critical value indicating that there is enough evidence to reject the null hypothesis. In return, this indicates that the model developed the this covariate combination is acceptable. A plot of the cumulative PDF and the EDF are given in Figure 5.4.

The test statistic, model parameters and log-likelihood value obtained with this covariate combination is given in Table 5.3.

Although this combinations provides an acceptable model fit, further investigation has to be done to determine whether another covariate combination provides better a model fit.

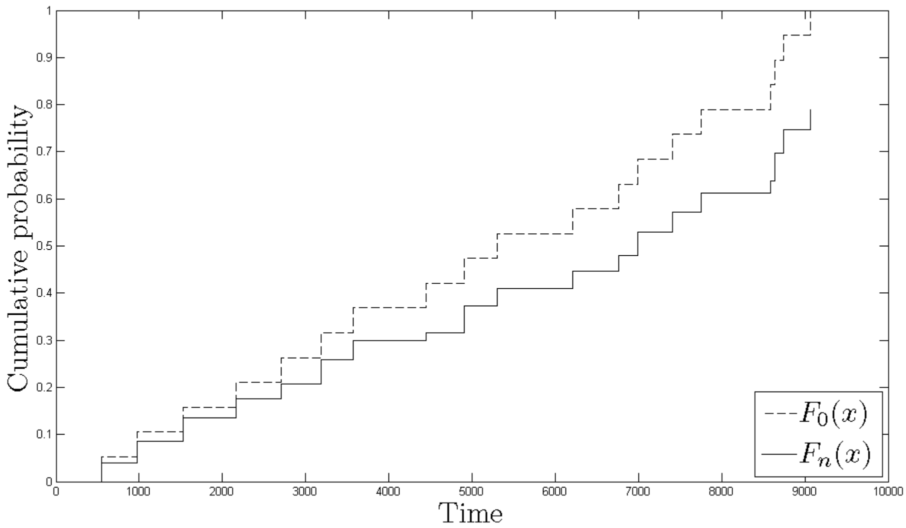

Figure 5.4: Covariate combination: Fe, boost pressure, left exhaust temperature and engine filter oil pressure (min).

Table 5.3: Covariate combination: Fe, boost pressure, left exhaust temperature and engine oil pressure.

D_n	0.2129
β	0.9617
η	6260.6033
γ_1	-1.2043
γ_2	0.0197
γ_3	-0.0049
γ_4	-4.6237
log-likelihood	6899.9

5.4.3.2 Covariate Combination 2

For the third combination the Fe, right after cooler temperature and engine coolant temperature covariates were used. For this particular scenario, D_n, is equal to 0.2105 indicating that there is enough evidence to reject the null hypothesis which means that this model is acceptable. A plot of the cumulative PDF and the EDF is given in Figure 5.5.

The test statistic, model parameters and log-likelihood value obtained with

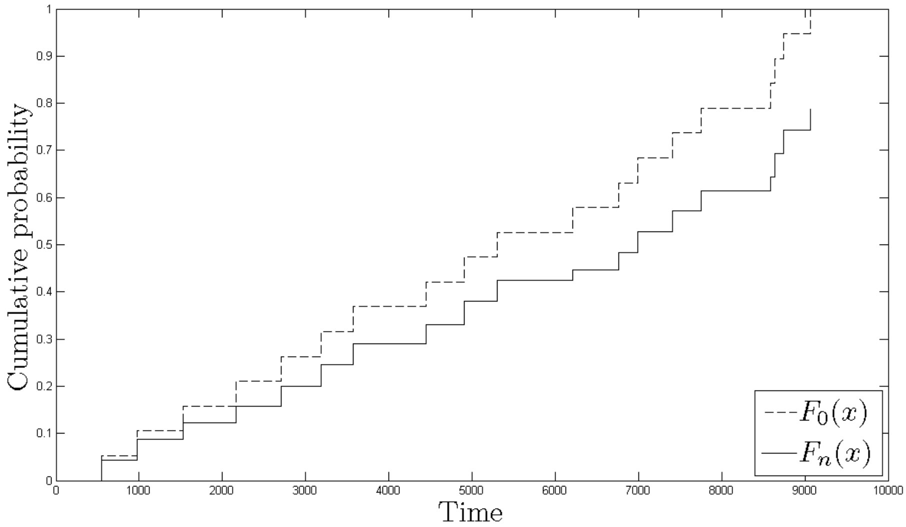

Figure 5.5: Covariate combination: Fe, right after cooler temperature and engine coolant temperature.

Table 5.4: Covariate combination: Fe, right after cooler temperature and engine coolant temperature.

D_n	0.2105
β	0.9449
η	1489.1816
γ_1	0.6044
γ_2	-0.0204
γ_3	0.0077
log-likelihood	34.883

this covariate combination is given in Table 5.4.

Table 5.5: Covariate combination: Fe, boost pressure and left exhaust temp.

D_n	0.1731
β	0.883
η	365.542
γ_1	0.394
γ_2	0.026
γ_3	-0.011
log-likelihood	34.8226

5.4.3.3 Covariate Combination 3

The final combination consisted of Fe, boost pressure and left exhaust temperature. The calculated test statistic, D_n, is equal to 01731 which meant that there is enough evidence to reject the null hypothesis indicating that the covariates used provide an acceptable model fit. A plot of the cumulative PDF and the EDF is given in Figure 5.6.

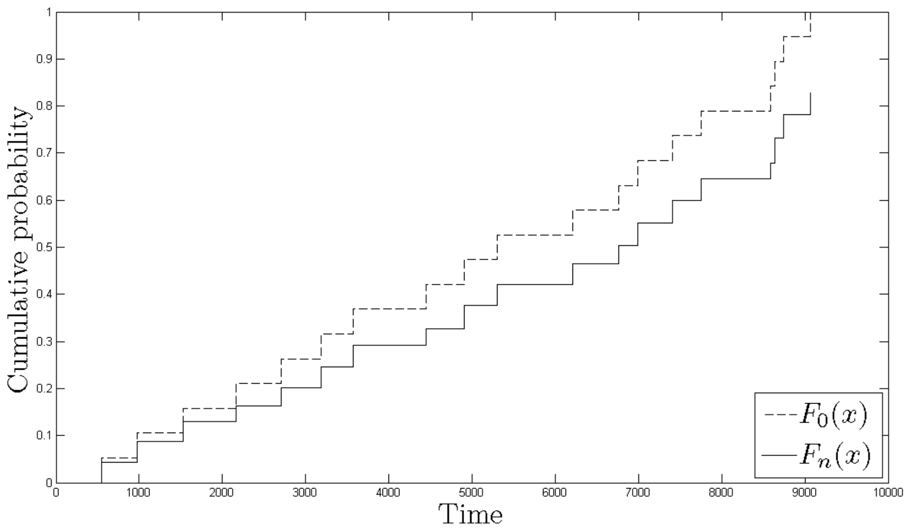

Figure 5.6: Covariate combination: Fe, boost pressure and left exhaust temp.

The test statistic, model parameters and log-likelihood value obtained with this covariate combination is given in Table 5.5.

5.4.3.4 Final Covariate Combination

In Table 5.6 the K-S test statistic results for the covariate combinations are tabulated. In each case the test statistic was found at the last measurement. This indicated that that the cumulative PDF and EDF tended to diverge. However, according to Equation 5.4.2, if the test statistic, D_n, is less than the critical value, $c_{nj\alpha}$, there is not enough evidence to reject the null hypothesis.

With a critical value of $c_{nj\alpha} = 0.301$, there is not enough evidence to reject the null hypothesis for any of the four covariate combinations. However, the combination with best model fit is the third combination. As a result, the third combination is selected to develop the final Weibull PHM for the thesis.

Table 5.6: Kolmogorov-Smirnoff values.

Covariate Combination	D_n
Fe, boost pressure, engine oil pressure (high) and (low).	0.2151
Fe, boost pressure, left exhaust temperature and engine oil pressure (low).	0.2129
Engine oil pressure (low) and right after cooler temperature.	0.2105
Fe, boost pressure and left exhaust temperature.	0.1731

5.4.4 Final Model

The K-S tests showed that the third covariate combination had the best model fit and is consequently used for the final model. The covariates in this combination was Fe (z_1), boost pressure (z_2) and left exhaust temperature (z_3). It can then be shown that the final Weibull PHM model is given by:

$$h(x, \overline{z(x)}) = \left(\frac{0.883}{365.542}\right)\left(\frac{x}{365.542}\right)^{0.883-1} \exp\left(z_1 \cdot 0.394 + z_2 \cdot 0.026 - z_3 \cdot 0.011\right) \tag{5.4.3}$$

With this model the decision models are constructed.

5.5 Model Validation Data

In Section 5.4.1 it was stated that seven of the eight CAT 793D haul trucks is used to develop a Weibull PHM model. The eighth haul truck's data is then used to

Table 5.7: Validation truck event's table.

#	C_i	X_i	$Z_1(X_1)$ [ppm]	$Z_2(X_2)$ [kPa]	$Z_3(X_3)$ [°C]
1	0	530	0.24	172.5	678
2	0	384	0.16	151	631
3	0	445	0.11	179.5	672
4	1	341	0.2259	185	651
5	0	445	0.24	182.5	662

validate the model. From this point forward this haul truck is referred to as the validation truck. This section discusses how the data of the validation truck was modified to be able to use it in the decision model.

The same data manipulation procedures, as discussed in Section 5.4.1, are applied to the validation truck. Five events were observed in the validation truck data, four of which was censored observations and one failure. These events are tabulated in Table 5.7.

Where $Z_1(X_1)$, $Z_2(X_2)$ and $Z_3(X_3)$ represent Fe, boost pressure and left exhaust temperature. In order to test the validity of the model it has to be applied to the validation truck data and predict the time of the events. The validation truck data is handled as follows:

(i.) Data up to a specified inspection point X_c is extracted from an event's history. This decision is based on the fact that oil samples are taken weekly, i.e. every 168 hours. For this reason it is assumed that it is most practical to specify data up to this point.

(ii.) A prediction of the data is then made into the future from the point of inspection, up to the point where the survivor function approaches.

Due to the fact that there is no data available onwards from the point of inspection, future covariate behaviour has to be extrapolated into the unknown. Procedures considered for future covariate behaviour prediction was keeping the covariate value constant and fitting a trend line to the data up to the time of the inspection.

Several attempts were undertaken where the covariate values were taken to be constant into the future. A problem with this method was that the reliability functions were totally erroneous due to the fact that RUL estimations were in most cases two orders too big. As a result, this procedure was not used for covariate future prediction.

The second procedure entails fitting a trend line to the data up to the 168^{th} hour and extrapolating the covariate values into the future. It was found that this assumption lead to acceptable reliability functions. After fitting different trend to the covariates, it was found that the best trend line for $Z_1(X_1)$ and $Z_2(X_2)$ were linear lines and second order polynomials for $Z_3(X_3)$. Figure 5.7 illustrates how $Z_2(X_2)$ was extrapolated into the future from the 168^{th} hour.

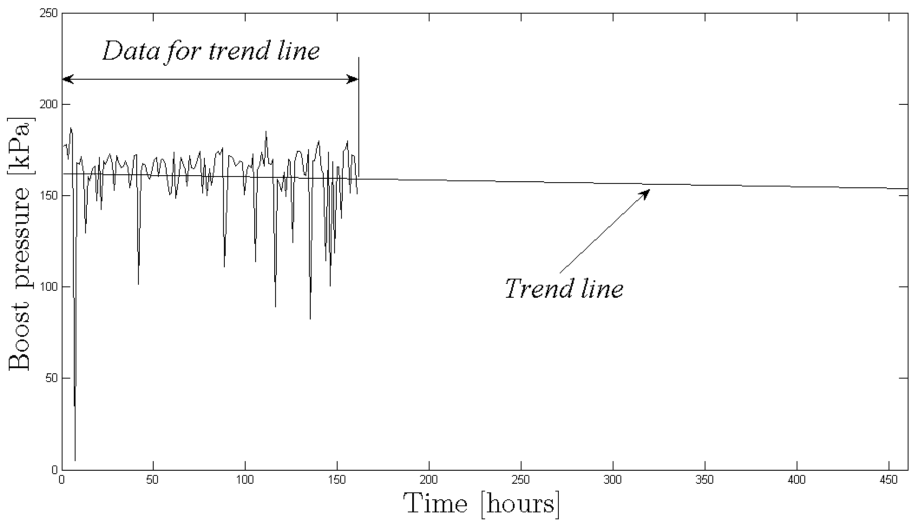

Figure 5.7: Extrapolation of boost pressure data into the future.

The same procedure is followed for $Z_1(X_1)$. The best trendline foudn for $Z_3(X_3)$ was second order polynomials to extrapolate the data as shown in Figure 5.8.

These procedures are followed whenever it is required to predict future covariate behaviour. This is particularly important for RUL and optimal cost function estimations with which the model is validated.

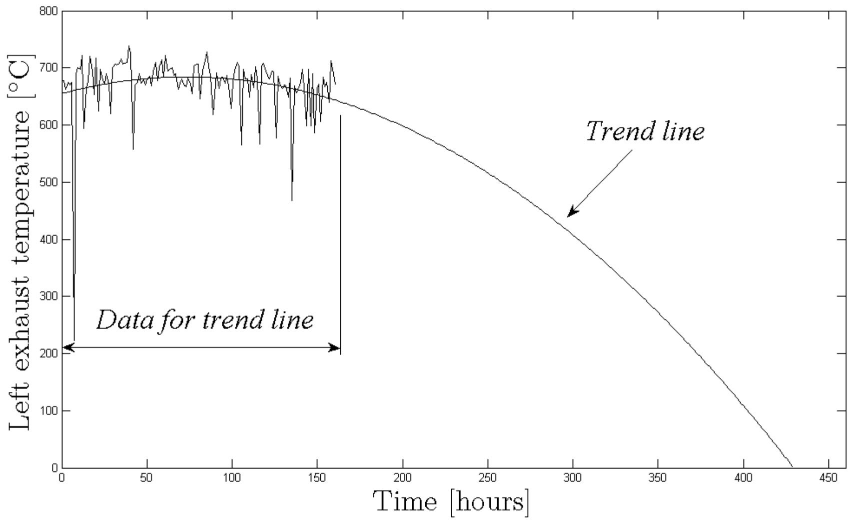

Figure 5.8: Extrapolation of left exhaust temperature data into the future.

5.5.1 Decision Model

In Chapter 3, three decision model approaches were introduced, namely: (a) Residual Useful Life and (b) long term cost optimization. The decision models are repeated here for convenience sake.

- Decision model 1: Perform maintenance according to the residual useful life estimates, i.e.

$$\underset{\sim}{X}_{r+1} - X_c/2 \leq x \tag{5.5.1}$$

  where $\underset{\sim}{X}_{r+1}$ is the lower confidence limit, X_c the inspection interval and x the current oil age in hours.

- Decision model 2: Perform maintenance according to the optimal point on the cost function.

$$x = X_p \tag{5.5.2}$$

  where x is the current point in time and X_p the recommended preventive replacement time.

- Decision model 3: Combine decision model 1 and 2 and perform maintenance whenever either of the two decision models suggests that maintenance has to be performed.

After each model is presented, the models are compared to establish which model produces the best event arrival time predictions. It is assumed that decision model predictions can be made when weekly oil samples are taken. For this reason, the decision models determine event arrival predictions on a weekly basis.

Due to the fact that oil samples are taken weekly, i.e. every 168 hours, predictions are made every 168 hours. Additional to the weekly predictions, predictions are also made at the time of the event itself.

5.5.2 Decision model 1: Residual Useful Life Estimates

The decision model for RUL estimates state that maintenance is done whenever the lower RUL confidence limit reduces to less than the inspection interval duration X_c, i.e. 168 hours. Note, predictions and estimates are used interchangeably in this chapter.

5.5.2.1 Event 1

The first event occurred in the 530^{th} hour. Figure 5.9 illustrates the three RUL estimates at hours 168, 338 and 530, respectively. At each estimate the confidence limits are illustrated by vertical lines. At the 168^{th} hour, the RUL estimate is E[X_{r+1}]=369 hours with upper and lower confidence limits of $\tilde{X}_{r+1} = 419$ and $\underset{\sim}{X}_{r+1} = 303$ hours, respectively.

According to Equation 5.5.1, $\underset{\sim}{X}_{r+1} - X_c/2 = 215$ and $x = 168$ which means that maintenance should not be performed due to the fact that $215 \nleq 168$. Please note, from this point forward the results are directly compared and not given separately.

At the 338^{th} hour, the RUL estimate is E[X_{r+1}]=1561 hours with confidence limits of $\tilde{X}_{r+1} = 2018$ and $\underset{\sim}{X}_{r+1} - X_c/2 = 753$ hours, respectively. The increased prediction might be an indication that the condition of the haul truck improved since the previous prediction was made. As a result, maintenance should not be performed due to the fact that $665 \nleq 338$. The first two estimates indicated that maintenance should not have been performed which, according to the actual event time, is correct.

The final prediction made at the 530^{th} hour produced a RUL estimate of E[X_{r+1}]=543 hours with confidence limits of 618 and $\underset{\sim}{X}_{r+1} - X_c/2 = 542$ hours, respectively. Due to the fact that $454 \leq 530$, it is suggested that maintenance

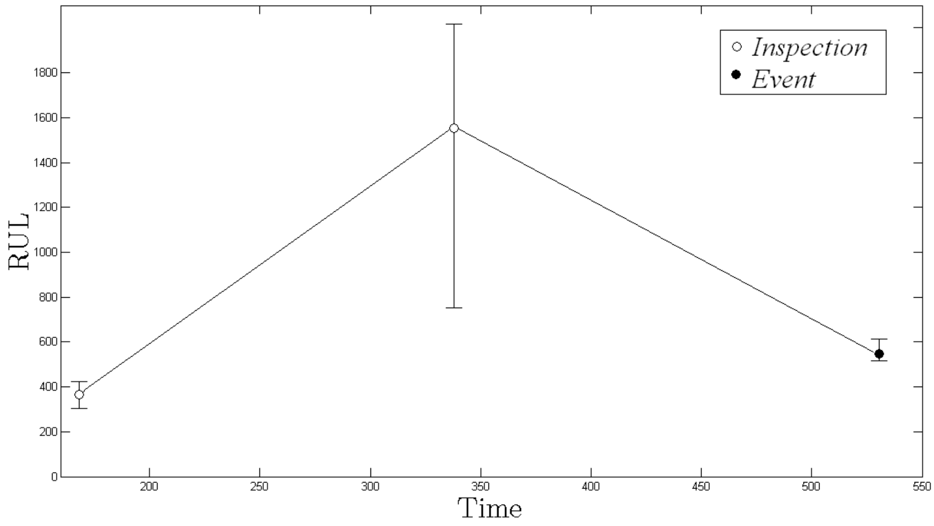

Figure 5.9: Validation truck event 1.

should be performed which coincides with the actual event arrival time and therefore an accurate prediction.

5.5.2.2 Event 2

The second event occurred in the 384^{th} hour. Figure 5.10 illustrates the three RUL estimates at hours 168, 338 and 384, respectively. At the 168^{th} hour, the RUL estimate is $\mathrm{E}[X_{r+1}]$=180 hours with confidence limits of $\tilde{X}_{r+1} = 207$ and $\underset{\sim}{X}_{r+1} = 170$ hours, respectively.

According to Equation 5.5.1, $\underset{\sim}{X}_{r+1} - X_c/2 = -86$ and $x = 168$ which indicates that maintenance has to be performed because $82 \leq 168$.

The prediction at the 338^{th} hour indicated that the RUL estimation is $\mathrm{E}[X_{r+1}]$=623 hours with confidence limits of $\tilde{X}_{r+1} = 648$ and $\underset{\sim}{X}_{r+1} = 605$ hours, respectively. Maintenance should not be performed due to the fact that $517 \nleq 338$. At the 384^{th} hour, the RUL estimate is $\mathrm{E}[X_{r+1}]$=820 hours with confidence limits of $\tilde{X}_{r+1} = 849$ and $\underset{\sim}{X}_{r+1} = 779$ hours, respectively. Maintenance should not be performed due to the fact that $691 \nleq 384$.

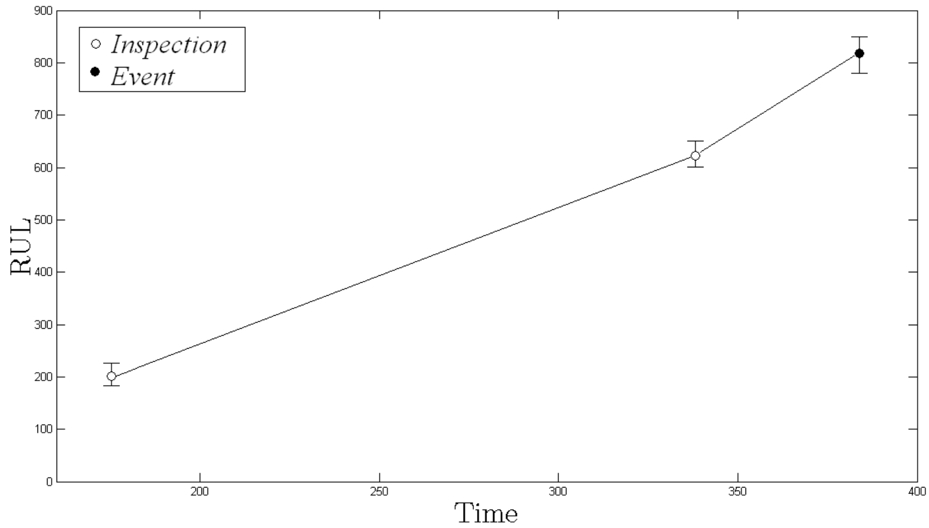

Figure 5.10: Validation truck event 2.

The first prediction proved to be conservative which meant that maintenance would have been done far too early. The last two estimations were found to be less conservative than the first prediction. However, maintenance decisions based on the estimates for the second event would not have been satisfactorily.

5.5.2.3 Event 3

The third event occurred in the 444^{th} hour. Figure 5.11 illustrates the three RUL estimates at hours 168, 338 and 444, respectively. At the 168^{th} hour, the RUL estimate is determined to be $\mathrm{E}[X_{r+1}]$=502 hours with confidence limits of $\tilde{X}_{r+1} = 580$ and $\underset{\sim}{X}_{r+1} = 384$ hours, respectively.

According to Equation 5.5.1, $\underset{\sim}{X}_{r+1} - X_c/2 = 296$ and $x = 168$ which meant that the decision model does not suggest that maintenance should be taken, i.e. $296 \nleq 168$.

The second prediction at the 338^{th} hour is $\mathrm{E}[X_{r+1}]$=879 hours with confidence limits of $\tilde{X}_{r+1} = 1039$ and $\underset{\sim}{X}_{r+1} = 604$ hours, respectively. Maintenance should not be performed at this point due to the fact that $516 \nleq 444$. At the 444^{th} hour, the RUL estimate is $\mathrm{E}[X_{r+1}]$=626 hours with confidence limits of $\tilde{X}_{r+1} = 680$ and $\underset{\sim}{X}_{r+1} = 537$ hours, respectively. Maintenance should also not be performed due to

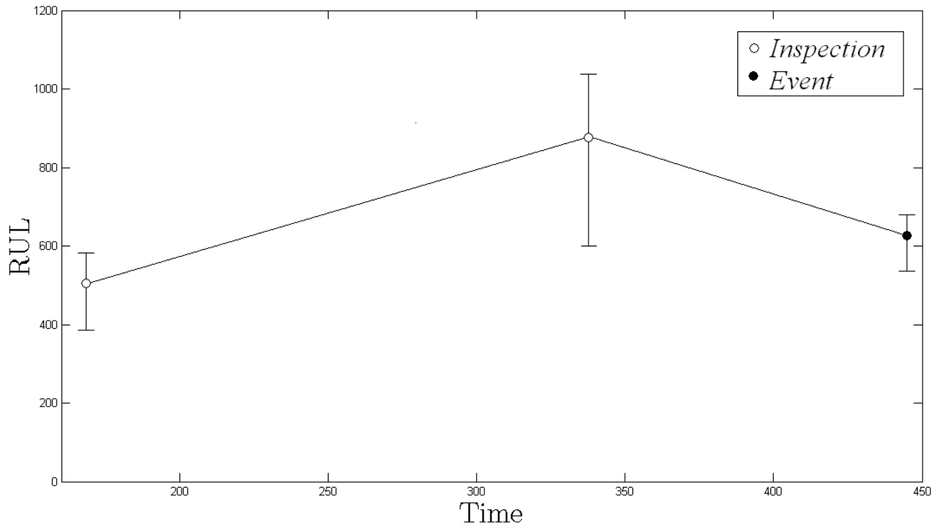

Figure 5.11: Validation truck event 3.

the fact that $449 \nleq 444$.

However, at the time of the event at the 444^{th} hour, the prediction that maintenance action should be taken at the 449^{th} hour is a good result even though it does not exactly represent reality. Estimates for the third event proved that the model had accurately predicted the the haul truck's third event and performing maintenance based on the decision model predictions would have been successful.

5.5.2.4 Event 4

The fourth event took place in the 341^{th} hour. Figure 5.11 illustrates three RUL estimates at hours 168, 337 and 341, respectively. The first prediction at the 168^{th} hour indicated that the RUK estimate E[X_{r+1}]=519 hours with confidence limits of $\tilde{X}_{r+1} = 636$ and $\underset{\sim}{X}_{r+1} = 353$ hours, respectively.

According to Equation 5.5.1, maintenance should not be performed at this point due to the fact that $\underset{\sim}{X}_{r+1} - X_c/2 = 265$ and $x = 168$ and $265 \nleq 168$.

The RUL prediction at the the 338^{th} hour is E[X_{r+1}]=447 hours with confidence limits of $\tilde{X}_{r+1} = 473$ and $\underset{\sim}{X}_{r+1} = 412$ hours, respectively. In this case maintenance should be performed due to the fact that $324 \leq 338$. The final prediction at the 341^{st} hour indicated that the RUL is E[X_{r+1}]=567 hours with confidence limits

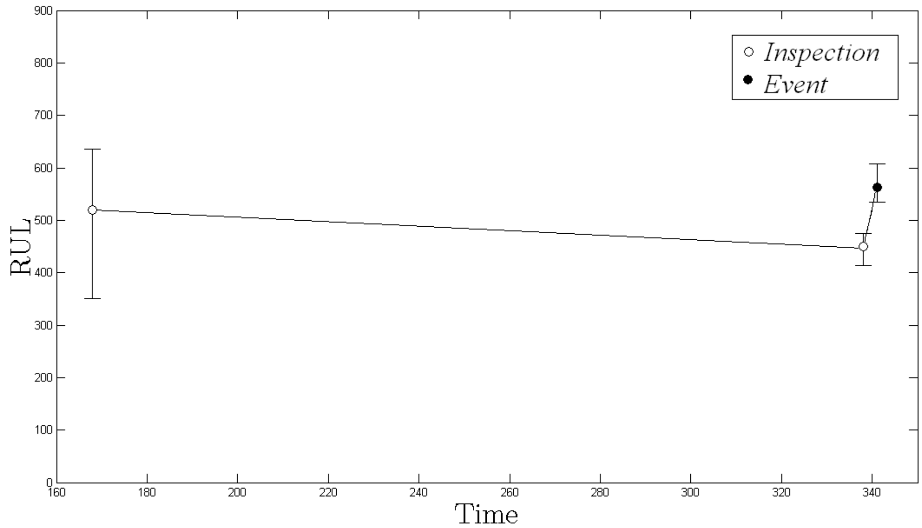

Figure 5.12: Validation truck event 4.

of $\tilde{X}_{r+1} = 618$ and $\underset{\sim}{X}_{r+1} = 554$ hours, respectively. Maintenance should not be performed due to the fact that $466 \nleq 341$.

The second estimate indicated that maintenance had to be performed in the 338^{th} hour. The difference between the second estimate and the actual event time is merely three hours. Therefore it can be assumed that the model accurately modelled the haul truck's fourth event and would have been a good decision to perform maintenance using the decision model predictions for this event.

5.5.2.5 Event 5

The final event occurred in the 445^{th} hour. Figure 9 illustrated that the RUL estimate is in the 445^{th} hour. At this point the RUL estimate is $\text{E}[X_{r+1}]$=641 hours with upper and lower confidence limits of $\tilde{X}_{r+1} = 675$ and $\underset{\sim}{X}_{r+1} = 539$ hours, respectively.

This meant that maintenance should not be performed due to the fact that $\underset{\sim}{X}_{r+1} - X_c/2 = 451$ and $x = 445$ and $451 \nleq 445$. However, the fact that the estimate and actual event time differ by only 6 hours indicated that the model accurately represented the failure behaviour the haul truck's fifth event and performing maintenance based on the predictions would have been satidfactorly.

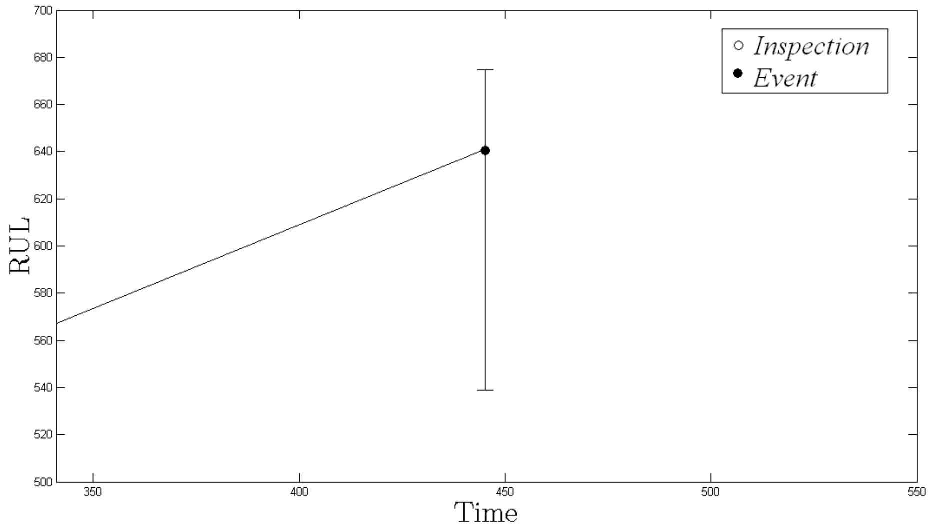

Figure 5.13: Validation truck event 5.

5.5.2.6 Conclusion

Estimations made for events one and four indicated that maintenance had to be performed. This would have been good decisions due to the fact that it was accurate predictions of the actual event arrival times. However, estimations made for events three and five indicated that maintenance had to be performed a few hours after the actual event arrival time. Due to the fact that these predictions were inaccurate by a few hours, it was assumed that it would have resulted in good maintenance decisions.

The overall accuracy of the estimates would have ensured that good decisions would have been made if the model was used to determine maintenance instants which in essence is the development of ACPs.

5.5.3 Decision model 2: Long Term Cost Optimization

The costs considered in this section are approximate maintenance costs. Due to company policies it is not possible to obtain exact maintenance costs. Costs were estimated to be $C_p = R65\ 000$ for preventive maintenance and $C_f = R1\ 600\ 000$ for failure maintenance. As with the RUL estimates, a cost function estimate was determined at each inspection interval and time of event. Cost function estimates found in this section were found using Equation 3.7.12 in Chapter 3.

5.5.3.1 Event 1

Optimal maintenance instances for the first event were determined at hours 168, 338 and 530, respectively. At the first inspection point it was indicated that the optimal maintenance instant was at the 225^{th} hour at a cost of R487.88/hour. The optimal maintenance instant calculated at the second inspection point was calculated to be at the 749^{th} hour at a cost of R252.89/hour.

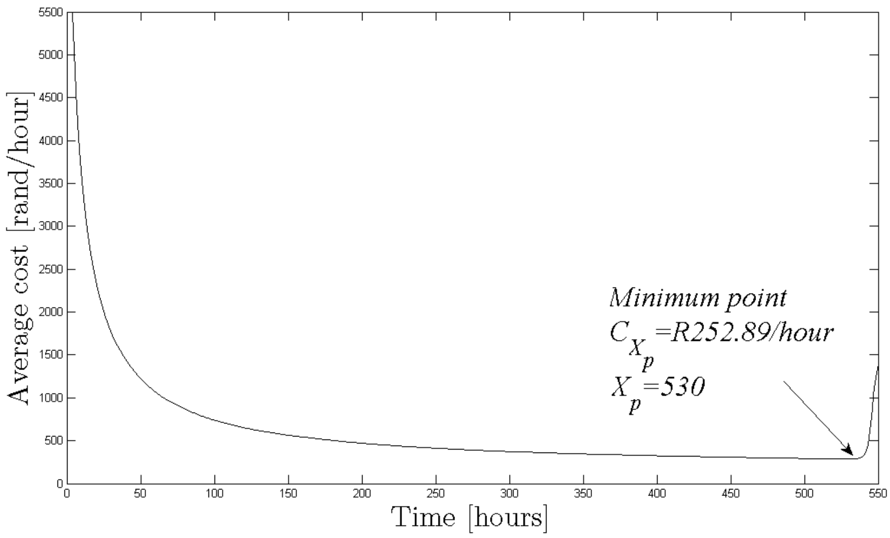

Figure 5.14: Long term cost optimization function for event 1.

At the final inspection point the optimal maintenance instant was calculated to be at the 530^{th} hour at a cost of 285.58/hour. A cost function plot for the last estimate is graphically illustrated in Figure 5.14. Here the minimum point is found where $x = 530$ with an average hourly cost of R258.58/hour. Graphical illustrations of the first two cost function's are illustrated in Figures B.1 and B.2.

The first two inspection point estimations did not indicate that maintenance action had to be taken. However, the third estimation indicated that maintenance action had to be taken at $x = 530$ hours.

5.5.3.2 Event 2

For the second event, optimal maintenance instances were determined at hours 168, 338 and 384, respectively. At the first inspection point the maintenance instant was calculated to be at the 560^{th} hour at a cost of R584.01/hour. As for the second inspection point, the optimal maintenance instant was calculated to be at the 577^{th} hour at a cost of R233.98/hour.

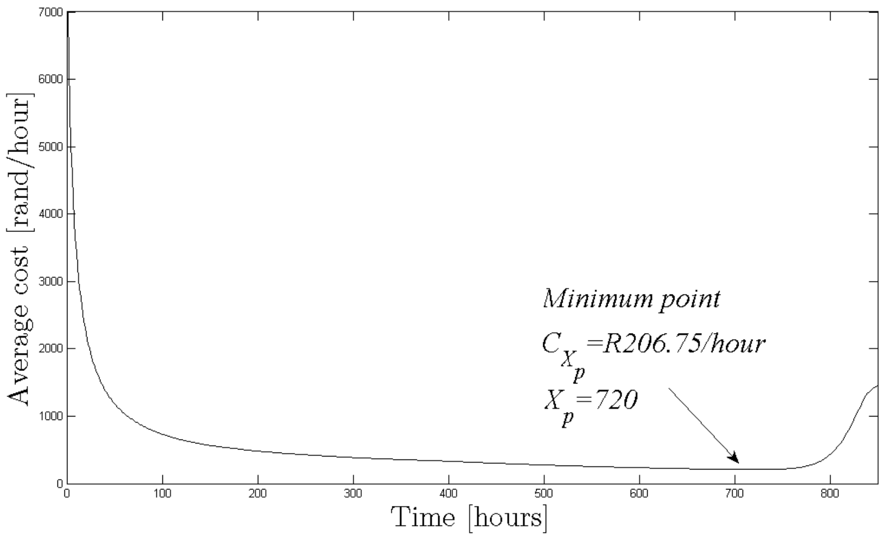

Figure 5.15: Long term cost optimization function for event 2.

The cost function estimate at the final inspection point were calculated to be at the 530^{th} hour at a cost of R206.75/hour. Figure 5.15 graphically illustrates the cost function at the final inspection point. It is clear to see that the minimum point is found where $x = 720$ and the average hourly cost is R206.75/hour.

Graphical illustrations of the first two inspection point cost function's are illustrated in Figures B.3 and B.4. These estimates indicated that no maintenance action had to be taken at the times of inspection.

5.5.3.3 Event 3

For the third event, optimal maintenance instances were determined at hours 168, 338 and 444, respectively. The optimal maintenance instant estimated at the first inspection point was calculated to be at the 293^{th} hour at a cost of R443.06/hour. At the second inspection point, the optimal maintenance instant was determined to be at the 506^{th} hour at a cost of R316.8/hour.

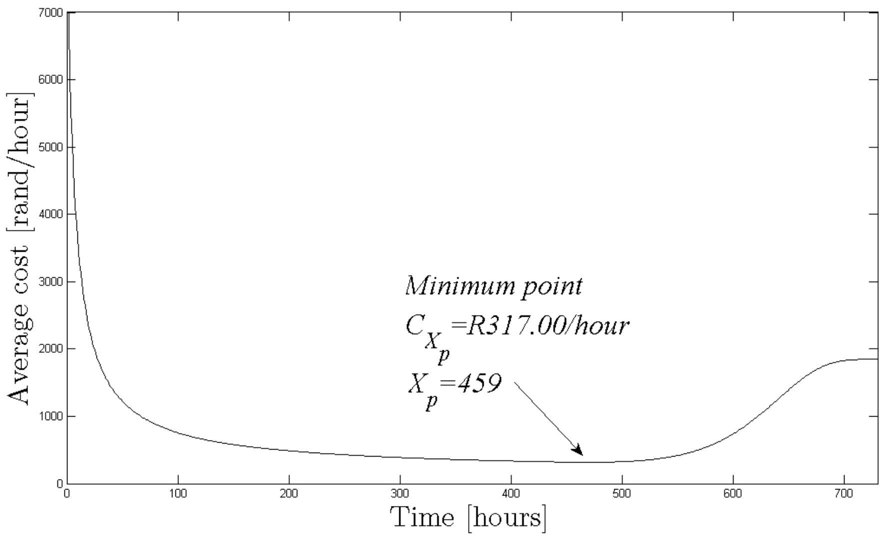

Figure 5.16: Long term cost optimization function for event 3.

The final inspection point indicated that the optimal maintenance inspection instance was found to be at the 459^{th} hour at a cost of R317/hour. A cost function plot for the estimate is graphically illustrated in Figure 5.16 where the minimum point is found at $x = 577$ where the average hourly cost is R317/hour. Graphical illustrations of the first two cost functions are illustrated in Figures B.5 and B.6.

These estimates indicated that no maintenance action had to be taken at the times of inspection.

5.5.3.4 Event 4

Optimal maintenance instances was calculated at hours 168, 338 and 341, respectively. At the first inspection point, the optimal maintenance instant was calculated to be at the 272^{th} hour at a cost of R418.39/hour. The optimal maintenace instant calculatd at the second inspection point was found to be at the 374^{th} hour at a average cost of R419.37/hour.

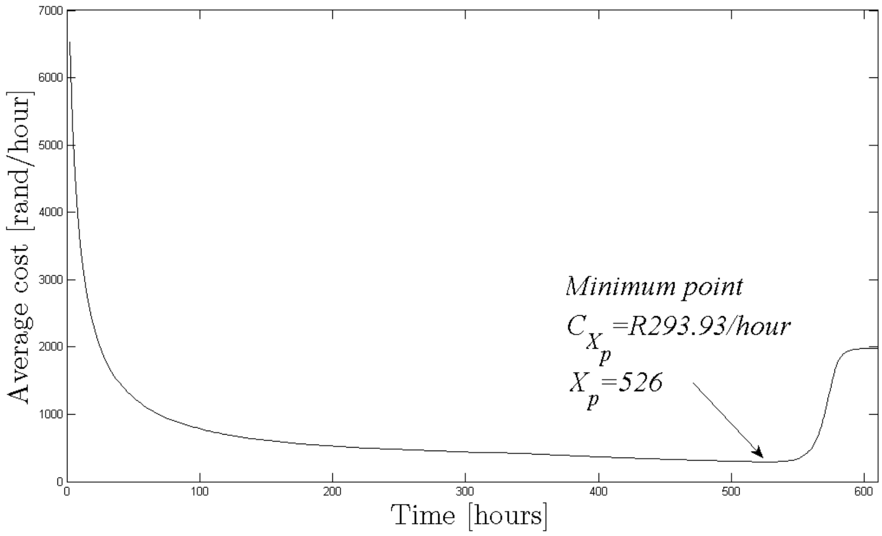

Figure 5.17: Long term cost optimization function for event 4.

Calculation of the optimal maintenance instance at the final inspection point was determined to be at the 526^{th} hour at a cost of R293.93/hour. The final inspection point cost function graphically illustrated in Figure 5.17 where the minimum point is found at $x = 577$ where the average hourly cost is R293.93/hour. Graphical illustrations of the first two inspection point cost functions are illustrated in Figures B.7 and B.8.

These estimates indicated that no maintenance action had to be taken at the times of inspection.

5.5.3.5 Event 5

Only one optimal maintenance instant calculation was done for the fifth event. It was found that the optimal instant was calculated to be at the 472^{nd} hour at a cost of R359.10/hour. The cost function is graphically illustrated in Figure 5.18 where the minimum point is found at $x = 472$ at an average cost of R359.10/hour. This estimate did not suggest that maintenance action had to be taken.

5.5.3.6 Conclusion

Although most of the results given in this section indicated that maintenance had to be performed later than the actual event arrival times, the overall accuracy of

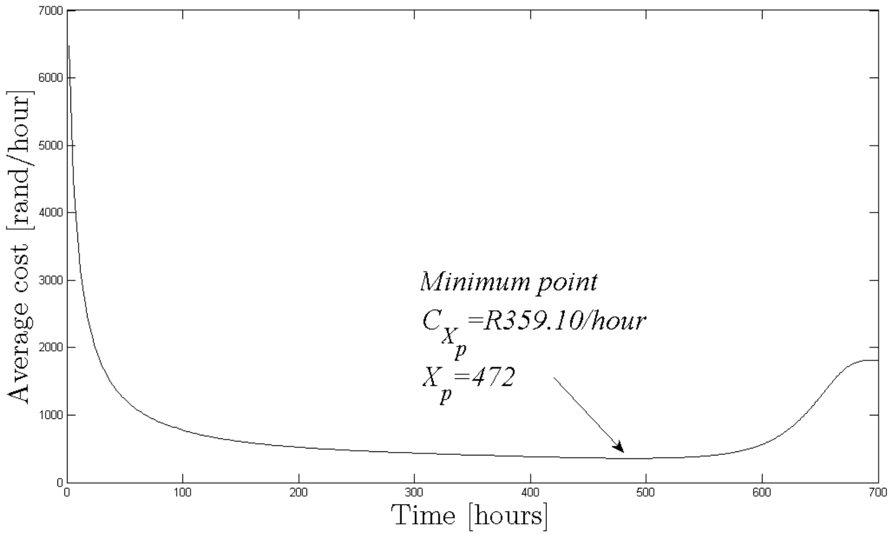

Figure 5.18: Long term cost optimization function for event 5.

the model to represent the failure behaviour of the haul trucks were satisfactory. If this decision model is used it would not prevent al the failures but it is a good representation of reality.

For this reason, the estimates determined in this section would also have ensured fairly accurate maintenance decisions would have been made if this maintenance decision model was used with the developed Weibull PHM.

5.5.4 Decision model 3: Combining decision model 1 and decision model 2.

This section combines results obtained with decision model 1 and decision model 2 in order to develop the third decision model. Results obtained at each inspection point for each respective event for both the first and second decision models are tabulated in Table 5.5.4. The third decision model adopts the earliest prediction of the other two decision models. For example, in Table 5.5.4 the first and second event arrival times of the first and second decision models are 215 and 225, respectively. In this case the third decision model would adopt the first decision model's result because 215≤225.

The next step is to determine which model is best suited for maintenance scheduling. For each inspection result a percentage error value is determined.

Percentage error values are calculated using the following expression:

$$\%error = abs(\frac{event\ time - model\ decision\ time}{event\ time}) \tag{5.5.3}$$

Calculating the percentage error values determines which decision model predicted the time of event most accurately. The percentage error illustrates the error of the decision model prediction based on the actual arrival time of the event. This method compares the decision models in a crude but effective way. The total for each error percentage column can be seen in Table 5.5.4.

The minimum total percentage error value indicates the best decision model for maintenance scheduling due to the fact that it exhibits the least deviation from predicting the correct event arrival times. Therefore the decision best suited for maintenance intervention scheduling is the second decision model.

Table 5.8: Criteria 1 and 2 combination.

Decision Model							
1			2			3	
$X_{r+1} - X_c/2$	MA	% Error	X_p	MA	% Error	MA	% Error
215	No	0.59434	225	No	0.575472	No	0.59434
327	Yes	0.383019	749	No	0.413208	No	0.383019
454	Yes	0.143396	530	Yes	0	Yes	0.143396
82	Yes	0.786458	560	No	0.458333	No	0.786458
517	No	0.346354	577	No	0.502604	No	0.346354
691	No	0.799479	720	No	0.875	No	0.799479
296	No	0.333333	293	No	0.34009	No	0.34009
324	Yes	0.27027	506	No	0.13964	No	0.27027
466	No	0.04955	459	No	0.033784	No	0.033784
265	No	0.222874	272	No	0.202346	No	0.222874
324	Yes	0.27027	374	No	0.157658	No	0.27027
466	No	0.04955	526	No	0.184685	No	0.04955
451	No	0.013483	472	No	0.060674	No	0.013483
	Total:	4.262		Total:	3.943	Total:	4.253

5.6 Model Validation

Figure 5.19 illustrates the different event estimations together with the actual event times.

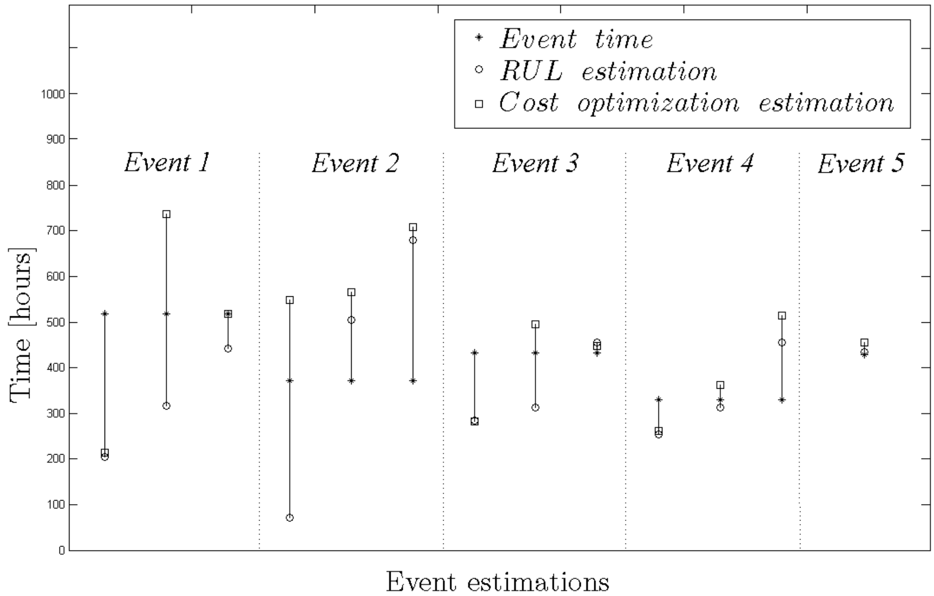

Figure 5.19: Decision model estimations compared to actual event times.

The first two estimations of the first event were found to be relatively inaccurate compared to the actual event arrival times. The problem was that data obtained for this event was minimal which meant that the model did not have enough information to accurately predict the arrival time of the first event at the first two inspections.

Another reason might be that the assumptions about future covariate behaviour could be too conservative due to the fact that four of the six estimates are earlier than the actual event time. However, the third estimation coincided with the failure time itself. This is due to the fact that sufficient data was available with which the model could accurately predict the event arrival time.

If the assumptions made about the future covariate behaviour are proved to be too conservative, the data indicates that the haul truck is in a "worse" state than it actually is which results in conservative estimations. If however the assumptions made about the future covariate behaviour are proven to exhibit a high level of risk, the data indicates that the truck is in a "healthier" state than it actually is which lead to estimations with a certain level of risk.

Estimations of the second event were also found to be slightly inaccurate. Further investigation indicated that several assumptions had to be made about the covariates. Seeing that five of the six estimations (RUL and cost optimization) were later than the actual failure times, assumptions regarding future covariate behaviour were exposed to allowing too high a level of risk.

Relative to the actual event times, the final three event estimates proved to be sufficiently acceptable. Estimates of the third event were rather conservative due to the fact that three of the estimates were earlier than the actual event times. As a consequence, assumptions regarding future covariate behaviour for this event proved too conservative. Final estimates of the fourth event indicated that the assumptions regarding future covariate behaviour were exposed to allowing too high a level of risk.

5.7 Conclusion

This chapter looked at the third item of the research methodology. It presented the development of the final Weibull PHM development and its application to a CAT 793D haul truck engine data. Results indicated that the accuracy of predicting future event arrival times proved to be accurate compared to the actual event arrival times. As a result, the final Weibull PHM is a good representation of the actual failure behaviour of the CAT 793D haul truck used for the case study.

The final Weibull PHM developed in this thesis, satisfactorily predicted event arrival times of a CAT 793D haul truck. Decision models were evaluated and it was concluded that the long term cost optimization decision model proved to be most successful and can be considered for ACP development.

Chapter 6

Conclusion and Recommendations

6.1 Overview

Chapter 1 indicated that a large amount of wasted maintenance activities are being performed. For this reason, industry is experiencing an increased interest in Physical Asset Management (PAM) to improve maintenance activities. The concept of PAM comprises of several Key Performance Areas (KPAs) such as Asset Care Plans (ACPs). These ACPs consist of many tactical maintenance strategies. Usage based Maintenance (UBM) bases its maintenance activities on historical failure data, not taking into account the condition of the item. Condition Based Maintenance (CBM) bases its maintenance activities on Condition Monitoring (CM) information, disregarding the historical failure data.

Each of these strategies take into account one type of data disregarding other data. For this reason, the problem undertaken in this thesis was to research models which take into account both types of data. Therefore, the objective of this study was to research the relevant literature in order to gain a detailed understanding of the models which can be used to incorporate both types of data.

Chapter 2 presented a comprehensive literature study which lead to Prognostics as a solution to the problem. It was then indicated that it might be possible to use regression models and it was therefore found necessary to investigate regression models.

Chapter 3 evaluated six regression models according to four criteria which reflected the applicability of each model to achieve the objectives of the thesis. These evaluations were compared revealing that the PHM is the most suitable model for this thesis.

In Chapter 4, two different forms of the PHM were discussed indicating that the fully-parametric Weibull PHM was found to be the most suitable form of the PHM. Parameter estimation of the fully-parametric form was done using an algorithm developed by Carstens *et al.* (2011). Several Goodness-of-fit models were introduced with which the model parameters could be tested. Lastly, three decision models were proposed with the third model being a combinations of the first and second decision model.

Chapter 5 presented the development of the final Weibull PHM and its application to a CAT 793D haul truck engine data. The Weibull PHM model is given by:

$$h(x, \overline{z(x)}) = \left(\frac{\beta}{\eta}\right)\left(\frac{x}{\eta}\right)^{\beta-1} \exp\left(\overline{z_i(x)} \cdot \bar{\gamma}\right) \tag{6.1.1}$$

Suitable data for the study was found at the open pit iron ore mine in the Northern Cape, Sishen. Validation of the Weibull PHM was done using the following procedure:

1. Fit the Weibull PHM to seven of the eight haul trucks' data.
2. Apply the developed model to the data obtained from the eighth truck to determine how accurate the estimates represent its failure behaviour.

Estimations obtained with the model indicated that the model was a relatively good representation of reality compared to the actual event arrival times. The three decision models were evaluated and it was concluded that the long term cost optimization decision model proved to be most successful and can be considered for ACP development assistance.

Two major benefits of the developed model are: a) The possibility of making relatively accurate estimations compared to the arrival times of events; and b) Reduced dependence on the expertise and experience of engineers and technicians for maintaining assets. In conclusion, from the obtained results it is concluded that the Weibull PHM is a good representation of reality.

Although the Weibull PHM provided satisfactorily results, there is still room for improvement. In Section 6.2 three recommendations are given with which the accuracy of the obtained result can be improved.

Findings from this study provide an exciting basis for the development of future Weibull PHMs that could result in huge maintenance cost savings and reduced failure occurrences.

6.2 Recommendations for Future Research

Experience gained through completing this thesis has led to three recommendations for future research:

i) Future covariate behaviour: The decision models applied in this thesis required that assumptions regarding future covariate behaviour had to be made. In Section 5.5 it was stated that the use of future covariate behaviour predictions elevate the importance of a more intricate understanding into the behaviour of these covariates. Consequently, it is vital that future research include comprehensive technique studies for predicting future covariate behaviour more accurately.

ii) Observed event composition: The majority of events in the data used were censored observations. The implication is that the developed model is essentially modelling current maintenance behaviour to determine preventive maintenance action. Due to the fact that the results obtained in this study accurately represented these observations means that the PHM has great potential for future research.

It is recommended that for future research, obtained data should contain a greater number of observed failures than censored observations. This recommendation only introduces the concept that better data could improve on the already satisfactorily result obtained in this research project.

iii) Data quality: The quality of data found in the South African industry remains a problem. As PAM gains popularity, greater commitment towards good quality data recordings will be practised. Despite the fact that the data used for this study was lacking quality, acceptable and interpretable results were obtained. It is therefore recommended that this thesis can be used to illustrate to industry the potential of failure data analysis.

These recommendations can be used to improve results obtained with future research projects.

List of References

Andersen, P. (1982). Testing goodness of fit of cox's regression and life model. *Biometrics*, pp. 67–77.

Anderson, T. and Darling, D. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, pp. 765–769.

Ascher, H. and Feingold, H. (1984*a*). *Repairable systems reliability: modeling, inference, misconceptions and their causes.* M. Dekker. ISBN 0824772768.

Ascher, H.E. and Feingold, H. (1984*b*). *Repairable systems reliability: Modeling, inference, misconceptions and their causes.* Marcel Dekker.

Balakrishnan, n. and Rao, C. (2004). *Handbook of Statistics 23.* Elsevier, North Holland.

Banjevic, D. and Jardine, A. (2006). Calculation of reliability function and remaining useful life for a markov failure time process. *IMA Journal of Management Mathematics*, vol. 17, no. 2, p. 115.

Banjevic, D., Jardine, A., Makis, V. and Ennis, M. (2001). A control-limit policy and software for condition-based maintenance optimization. *INFOR-OTTAWA-*, vol. 39, no. 1, pp. 32–50.

Barnard, G. (1963). Some logical aspects of the fiducial argument. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 111–114.

Bekker, J. (2011). Simulation notes. *Stellenbosch University Industrial Department.*

Bennett, S. (1983*a*). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, vol. 2, no. 2, pp. 273–277.

Bennett, S. (1983*b*). Log-logistic regression models for survival data. *Applied Statistics*, pp. 165–171.

Bonabeau, E., Dorigo, M. and Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems.* 1. Oxford University Press, USA.

Breslow, N. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, pp. 45–57.

Caesarendra, W., Niu, G. and Yang, B. (2010). Machine condition prognosis based on sequential monte carlo method. *Expert Systems with Applications*, vol. 37, no. 3, pp. 2412–2420.

Carstens, W., Vlok, P. and Visser, T. (2011). Estimation of multivariate proportional hazards model parameters with artificial bee colony optimization. *ORION*.

Cheng, S., Wei, L. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, vol. 82, no. 4, p. 835.

Cheung, Y.B., Yip, P.S.F. and Karlberg, J.P.E. (2001). Fetal growth, early postnatal growth and motor development in pakistani infants. *International Journal of Epidemiology*, , no. 30, pp. 66–74.

Chinnam, R. and Baruah, P. (2003). Autonomous diagnostics and prognostics through competitive learning driven HMM-based clustering. In: *Neural Networks, 2003. Proceedings of the International Joint Conference on*, vol. 4, pp. 2466–2471. IEEE. ISBN 0780378989. ISSN 1098-7576.

Chinnam, R. and Baruah, P. (2004). A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems. *International Journal of Materials and Product Technology*, vol. 20, no. 1, pp. 166–179.

Ciampi, A. and Etezadi-Amoli, J. (1985). A general model for testing the proportional hazards and the accelerated failure time hypotheses in the analysis of censored survival data with covariates. *Communications in Statistics-Theory and Methods*, vol. 14, no. 3, pp. 651–667.

Coetzee, J. (1997). The role of NHPP models in the practical analysis of maintenance failure data. *Reliability Engineering & System Safety*, vol. 56, no. 2, pp. 161–168. ISSN 0951-8320.

Cox, D. (1972*a*). Regression models and life tables. *Journal of the Royal Statistical Society Series*, vol. 1, p. 187 220. ISSN 21.

Cox, D. (1972*b*). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220.

Cox, D. (1975). Partial likelihood. *Biometrika*, vol. 62, no. 2, pp. 269–276.

Cox, D. and Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 248–275.

Dale, C. (1985). Application of the proportional hazards model in the reliability field. *Reliability Engineering*, vol. 10, no. 1, pp. 1–14. ISSN 0143-8174.

Deloux, E., Castanier, B. and Bérenguer, C. (2009). Predictive maintenance policy for a gradually deteriorating system subject to stress. *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 418–431. ISSN 0951-8320.

Elsayed, E., Liao, H. and Wang, X. (2006). An extended linear hazard regression model with application to time-dependent dielectric breakdown of thermal oxides. *IIE Transactions*, vol. 38, no. 4, pp. 329–340.

Endrenyi, J., Aboresheid, S., Allan, R., Anders, G., Asgarpoor, S., Billinton, R., Chowdhury, N., Dialynas, E., Fipper, M., Fletcher, R. *et al.* (2001). The present status of maintenance strategies and the impact of maintenance on reliability. *Power Systems, IEEE Transactions on*, vol. 16, no. 4, pp. 638–646. ISSN 0885-8950.

Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics*, pp. 181–192.

Feng, D., Xuefeng, H. and Jiyong, C. (2009). Reliability assessment based on equipment condition vibration feature using proportional hazards model. *Journal of Mechanical Engineering*, p. 12.

Fogel, G. (). Pas55 delegate material. Asset Management Course at Stellenbosch University in Conjunction with Pragma.

Ghasemi, A., Yacout, S. and Ouali, M. (2009). Parameter estimation for condition based maintenance with proportional hazard model.

Goldman, J. (1967). Stochastic point processes: limit theorems. *The Annals of Mathematical Statistics*, vol. 38, no. 3, pp. 771–779.

Goode, K., Moore, J. and Roylance, B. (2000). Plant machinery working life prediction method utilizing reliability and condition-monitoring data. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 214, no. 2, pp. 109–122.

Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P. and Sun, Y. (2010). A review on reliability models with covariates.

Heiman, G. (2010). *Basic statistics for the behavioral sciences*. Wadsworth Pub Co.

Heng, A. (2000). Intelligent Prognostics of Machinery Health Using Suspended Condition Monitoring Data.

Heng, A., Zhang, S., Tan, A. and Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739. ISSN 0888-3270.

Isermann, R. (2005). Model-based fault-detection and diagnosis-status and applications. *Annual Reviews in control*, vol. 29, no. 1, pp. 71–85. ISSN 1367-5788.

Jardine, A. (1973). *Maintenance, Replacement and Reliability.* The Copp Clark Publishing Company.

Jardine, A., Anderson, P. and Mann, D. (1987). Application of the weibull proportional hazards model to aircraft and marine engine failure data. *Quality and reliability engineering international*, vol. 3, no. 2, pp. 77–82.

Jardine, A., Banjevic, D., Wiseman, M., Buck, S. and Joseph, T. (2001). Optimizing a mine haul truck wheel motors condition monitoring program use of proportional hazards modeling. *Journal of Quality in Maintenance Engineering*, vol. 7, no. 4, pp. 286–302.

Jardine, A., Lin, D. and Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510. ISSN 0888-3270.

Jiang, S., Landers, T. and Reed Rhoads, T. (2006). Proportional intensity models robustness with overhaul intervals. *Quality and Reliability Engineering International*, vol. 22, no. 3, pp. 251–263.

Kacprzynski, G., Sarlashkar, A., Roemer, M., Hess, A. and Hardman, B. (2004). Predicting remaining life by fusing the physics of failure modeling with diagnostics. *JOM Journal of the Minerals, Metals and Materials Society*, vol. 56, no. 3, pp. 29–35.

Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Fialure Time Data.* John Wiley and Sons, New York.

Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization. *Techn. Rep. TR06, Erciyes Univ. Press, Erciyes*.

Karaboga, D. and Akay, B. (2009). A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108–132.

Kelly, A. and Eastburn, K. (1982). Terotechnology. a modern approach to plant engineering. *Physical Science, Measurement and Instrumentation, Management and Education-Reviews, IEE Proceedings A*, vol. 129, no. 2, pp. 131–136.

Kim, Y., Rizzoni, G. and Utkin, V. (2001). Developing a fault tolerant power-train control system by integrating design of control and diagnostics. *International Journal of Robust and Nonlinear Control*, vol. 11, no. 11, pp. 1095–1114. ISSN 1099-1239.

Kirmani, S. and Gupta, R. (2001). On the proportional odds model in survival analysis. *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 2, pp. 203–216. ISSN 0020-3157.

Klein, J. and Moeschberger, M. (2003). *Survival analysis: techniques for censored and truncated data*. Springer Verlag.

Krishnaiah, P.R. and Sen, P.K. (1984). *Handbook of Statistics, Nonparametric Methods*, vol. 4. North Holland, Amsterdam.

Kvaloy, J. and Lindqvist, B. (1998). Ttt-based tests for trend in repairable systems data. *Reliability Engineering & System Safety*, vol. 60, no. 1, pp. 13–28.

Kwan, C., Zhang, X., Xu, R. and Haynes, L. (2003). A novel approach to fault diagnostics and prognostics. In: *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, vol. 1, pp. 604–609. IEEE. ISBN 0780377362. ISSN 1050-4729.

Li, Y., Kurfess, T. and Liang, S. (2000). Stochastic prognostics for rolling element bearings. *Mechanical Systems and Signal Processing*, vol. 14, no. 5, pp. 747–762.

Lin, D., Oakes, D. and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, vol. 85, no. 2, p. 289.

Louit, D., Pascual, R. and Jardine, A. (2009). A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data. *Reliability Engineering & System Safety*, vol. 94, no. 10, pp. 1618–1628. ISSN 0951-8320.

Love, C. and Guo, R. (1991). Application of weibull proportional hazards modelling to bad-as-old failure data. *Quality and reliability engineering international*, vol. 7, no. 3, pp. 149–157.

Lu, W. and Zhang, H. (2007). Variable selection for proportional odds model. *Statistics in medicine*, vol. 26, no. 20, pp. 3771–3781.

Lugtigheid, D., Jardine, A. and Jiang, X. (2007). Optimizing the performance of a repairable system under a maintenance and repair contract. *Quality and Reliability Engineering International*, vol. 23, no. 8, pp. 943–960.

Ma, J. and Li, J. (1995). Detection of localised defects in rolling element bearings via composite hypothesis test. *Mechanical Systems and Signal Processing*, vol. 9, no. 1, pp. 63–75. ISSN 0888-3270.

Madu, C. (2000). Competing through maintenance strategies. *International Journal of Quality & Reliability Management*, vol. 17, no. 9, pp. 937–949. ISSN 0265-671X.

Makis, V. and Jardine, A. (1992). Optimal replacement in the proportional hazards model. *INFOR*, vol. 30, no. 1, pp. 172–183.

Márquez, A., de León, P., Fernández, J., Márquez, C. and Campos, M. (2009). The maintenance management framework: A practical view to maintenance management. *Journal of Quality in Maintenance Engineering*, vol. 15, no. 2, pp. 167–178. ISSN 1355-2511.

Martinussen, T. and Scheike, T. (2006). *Dynamic Regression Models for Survival Data*. Springer Science+Business Media, Inc., New York.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 42, no. 2, pp. 109–142. ISSN 0035-9246.

Mechefske, C. (1998). Objective machinery fault diagnosis using fuzzy logic. *Mechanical systems and signal processing*, vol. 12, no. 6, pp. 855–862. ISSN 0888-3270.

Milbrodt, H. and Strasser, H. (1990). On the asymptotic power of the two-sided kolmogorov-smirnov test. *Journal of Statistical Planning and Inference*, vol. 26, no. 1, pp. 1–23.

Miller, R., Gong, G. and Muñoz, A. (1981). *Survival analysis*. Wiley New York.

Mobley, R. (2002). *An introduction to predictive maintenance*. Butterworth-Heinemann.

Moubray, J. (1997). *Reliability-centered maintenance*. Industrial Press Inc. ISBN 0831131462.

Murphy, S., Rossini, A. and Van der Vaart, A. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, pp. 968–976.

Nachlas, J. (2005). *Reliability engineering: probabilistic models and maintenance methods*, vol. 190. CRC Press.

Newby, M. (1988). Accelerated failure time models for reliability data analysis* 1. *Reliability Engineering & System Safety*, vol. 20, no. 3, pp. 187–197. ISSN 0951-8320.

Newby, M. (2008). Monitoring and maintenance of spares and one shot devices. *Reliability Engineering & System Safety*, vol. 93, no. 4, pp. 588–594. ISSN 0951-8320.

Newby, M. and Dagg, R. (). Optimal inspection and maintenance for stochastically deteriorating systems. *Preprint of Norwegian Institute of Technology*.

Nilsson, J. and Bertling, L. (2007). Maintenance management of wind power systems using condition monitoring systemsLife cycle cost analysis for two case studies. *Energy Conversion, IEEE Transactions on*, vol. 22, no. 1, pp. 223–229. ISSN 0885-8969.

Orbe, J., Ferreira, E. and Núñez-Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in medicine*, vol. 21, no. 22, pp. 3493–3510. ISSN 1097-0258.

Pampuri, S., Schirru, A., De Luca, C. and De Nicolao, G. (2011). Proportional hazard model with 1 penalization applied to predictive maintenance in semiconductor manufacturing. In: *Automation Science and Engineering (CASE), 2011 IEEE Conference on*, pp. 250–255. IEEE.

Paz, N. and Leigh, W. (1994). Maintenance scheduling: issues, results and research needs. *International Journal of Operations & Production Management*, vol. 14, no. 8, pp. 47–69. ISSN 0144-3577.

Percy, D. and Alkali, B. (2006). Generalized proportional intensities models for repairable systems. *IMA Journal of Management Mathematics*, vol. 17, no. 2, p. 171.

Percy, D. and Alkali, B. (2007). Scheduling preventive maintenance for oil pumps using generalized proportional intensities models. *International Transactions in Operational Research*, vol. 14, no. 6, pp. 547–563.

Peto, R. (1972). Contribution to discussion of paper by dr cox. *Journal of the Royal Statistical Society, Series B*, vol. 34, pp. 205–207.

Pham, H. (2006). *Springer handbook of engineering statistics*. Springer Verlag.

Pijnenburg, M. (1991). Additive hazards models in repairable systems reliability. *Reliability Engineering & System Safety*, vol. 31, no. 3, pp. 369–390. ISSN 0951-8320.

Pike, M. (1966). A method of analysis of a certain class of experiments in carcinogenesis. *Biometrics*, vol. 22, no. 1, pp. 142–161.

Pintelon, L. and Gelders, L. (1992). Maintenance management decision making. *European Journal of Operational Research*, vol. 58, no. 3, pp. 301–317. ISSN 0377-2217.

Qi, J. (2009). *Comparison of proportional hazards and accelerated failure time models*. Ph.D. thesis.

Qiu, J., Seth, B., Liang, S. and Zhang, C. (2002). Damage mechanics approach for bearing lifetime prognostics. *Mechanical systems and signal processing*, vol. 16, no. 5, pp. 817–829.

Ray, A. and Tangirala, S. (1996). Stochastic modeling of fatigue crack dynamics for on-line failure prognostics. *Control Systems Technology, IEEE Transactions on*, vol. 4, no. 4, pp. 443–451.

Samrout, M., Chātelet, E., Kouta, R. and Chebbo, N. (2009). Optimization of maintenance policy using the proportional hazard model. *Reliability Engineering & System Safety*, vol. 94, no. 1, pp. 44–52.

Seetharaman, P. and Chintagunta, P. (2003). The Proportional Hazard Model for Purchase Timing. *Journal of Business and Economic Statistics*, vol. 21, no. 3, pp. 368–382. ISSN 0735-0015.

Shyur, H., Elsayed, E. and Luxhøj, J. (1999). A general model for accelerated life testing with time-dependent covariates. *Naval Research Logistics*, vol. 46, no. 3, pp. 303–321. ISSN 1520-6750.

Sorsa, T., Koivo, H. and Koivisto, H. (1991). Neural networks in process fault diagnosis. *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 21, no. 4, pp. 815–825. ISSN 0018-9472.

Swanson, L. (2001). Linking maintenance strategies to performance. *International Journal of Production Economics*, vol. 70, no. 3, pp. 237–244. ISSN 0925-5273.

Tableman, M. and Kim, J. (2004). *Survival analysis using S: analysis of time-to-event data*. CRC Press.

Therneau, T. and Grambsch, P. (2000). *Modeling survival data: extending the Cox model*. Springer Verlag.

Tian, Z. and Liao, H. (2011). Condition based maintenance optimization for multi-component systems using proportional hazards model. *Reliability Engineering & System Safety*.

Venkatasubramanian, V., Rengaswamy, R. and Kavuri, S. (2003*a*). A review of process fault detection and diagnosis:: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 313–326. ISSN 0098-1354.

Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. (2003*b*). A review of process fault detection and diagnosis:: Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 293–311. ISSN 0098-1354.

Vlok, P. (1999). *Vibration Covariate Regression Analysis of Failure Time Data with the Proportional Hazards Model*. Ph.D. thesis, Department of Mechanical and Aeronautical Engineering, University of Pretoria.

Vlok, P. (2001). Dynamic Residual Life Estimation of Industrial Equipment based on Failure Intensity Proportions.

Vlok, P., Coetzee, J., Banjevic, D., Jardine, A. and Makis, V. (2002). Optimal component replacement decisions using vibration monitoring and the proportional-hazards model. *Journal of the operational research society*, vol. 53, no. 2, pp. 193–202.

Vlok, P., Wnek, M. and Zygmunt, M. (2004). Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. *Mechanical systems and signal processing*, vol. 18, no. 4, pp. 833–847.

Wang, L., Chu, J. and Wu, J. (2007). Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. *International Journal of Production Economics*, vol. 107, no. 1, pp. 151–163. ISSN 0925-5273.

Wang, P. and Vachtsevanos, G. (2001). Fault prognostics using dynamic wavelet neural networks. *AI EDAM*, vol. 15, no. 04, pp. 349–365.

Wang, W. (2007). A two-stage prognosis model in condition based maintenance. *European journal of operational research*, vol. 182, no. 3, pp. 1177–1187.

Wang, W., Golnaraghi, M. and Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, vol. 18, no. 4, pp. 813–831.

Wang, W., Scarf, P. and Smith, M. (2000). On the application of a model of condition-based maintenance. *The Journal of the Operational Research Society*, vol. 51, no. 11, pp. 1218–1227. ISSN 0160-5682.

Wenping, Z. and Akay, B. (2009). A Clustering Approach Using Cooperative Artificial Bee Colony Algorithm. *Hindawi Publishing Corporation*. ISSN 1026-0226.

Wong, E., Jefferis, T. and Montgomery, N. (2010). Proportional hazards modeling of engine failures in military vehicles. *Journal of Quality in Maintenance Engineering*, vol. 16, no. 2, pp. 144–155.

Wu, X. and Ryan, S. (2011). Optimal replacement in the proportional hazards model with semi-markovian covariate process and continuous monitoring. *Reliability, IEEE Transactions on*, vol. 60, no. 3, pp. 580–589.

Yang, S. and Prentice, R. (1999). Semiparametric Inference in the Proportional Odds Regression Model. *Journal of the American Statistical Association*, vol. 94, no. 445, pp. 125–126.

You, M., Li, L., Meng, G. and Ni, J. (2010). Two-zone proportional hazard model for equipment remaining useful life prediction. *Journal of Manufacturing Science and Engineering*, vol. 132, p. 041008.

You, M., Liu, F. and Meng, G. (2011). Proportional hazards model for reliability analysis of solder joints under various drop-impact and vibration conditions. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*.

Zheng, Y., Fang, H. and Wang, H. (2006). Takagi-Sugeno fuzzy-model-based fault detection for networked control systems with Markov delays. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, no. 4, pp. 924–929. ISSN 1083-4419.

Appendices

Data

On-board data and oil analysis data is combined in one dataset and given in Table 1. The on-board data measurements and oil analysis data taken into consideration is given. Here # indicates the specific truck, X the time and C_i the type of failure. $C_i = 1$ indicates a failure and $C_i = 0$ indicates a censored observation. Covariates listed in Table 1 and Table 2 are:

1. Air filter pressure [kPa]
2. Boost pressure [kPa]
3. Engine coolant temperature [°C]
4. Engine oil filter pressure [kPa]
5. Engine oil pressure (high) [kPa]
6. Engine oil pressure (low) [kPa]
7. Left exhaust temperature [°C]
8. Right exhaust temperature [°C]
9. Right after cooler temperature [°C]
10. Fe [ppm]
11. Si [ppm]
12. Na [ppm]
13. Pb [ppm]
14. Cu [ppm]
15. V100

Table 2 indicates the covariates for haul truck number one from approximately the 75^{th} hour to the 98^{th} hour. In total data for a period of six months were obtained.

Table 1: Event table.

#	Haul truck	X_i	C_i	Air Fltr [kPa]	Boost Pres [kPa]	Eng Cool Temp [°C]	Eng Oil Fltr High RPM & Cool>70 [kPa]	Eng Oil Pres High & Cool>74 [kPa]	Eng Oil Pres Low & Cool>74 [kPa]	Lt Exh Temp [°C]	Rt Aftrclr Temp [°C]	Rt Exh Temp [°C]	Fe [ppm]	Si [ppm]	Na [ppm]	V100	Pb [ppm]	Cu [ppm]
1	1	552.00	0	7.50	183.50	80	21.00	433.50	352.00	607	60	610	38.00	3.00	8.00	13.12	3.00	2.00
2	2	418.32	0	4.50	164.00	81	24.50	468.00	383.50	591	61	673	31.17	3.00	9.81	12.31	1.81	3.81
3	2	561.00	0	4.00	170.00	74	20.50	477.50	390.50	515	59	588	40.00	4.00	8.00	12.53	2.00	3.00
4	3	634.67	0	5.50	199.00	81	21.50	429.00	356.50	623	60	610	59.22	2.22	9.00	12.26	2.78	1.78
5	3	538.00	0	6.00	191.50	81	21.00	430.00	359.00	619	61	607	75.00	4.00	8.00	13.20	4.00	2.00
6	5	486.42	0	4.50	176.50	83	29.00	444.50	353.50	633	61	637	35.88	3.12	339.73	13.33	1.12	22.31
7	5	385.00	0	4.50	198.00	65	20.50	461.50	337.00	636	61	629	27.00	3.00	11.00	13.25	1.00	14.00
8	5	872.00	0	4.50	157.50	86	25.00	445.00	328.00	604	60	602	33.00	3.00	8.00	13.31	1.00	11.00
9	6	454.00	0	4.50	186.00	78	28.00	459.00	356.00	640	61	629	22.00	3.00	10.00	12.62	2.00	77.00
10	6	399.00	0	3.50	88.00	80	22.50	479.00	385.00	555	57	552	13.00	3.33	9.00	13.46	0.33	27.00
11	6	904.00	1	5.00	194.50	79	24.00	468.00	381.50	608	61	608	22.00	3.00	9.00	12.94	2.00	110.00
12	7	562.00	0	4.50	173.50	79	20.00	463.00	380.00	607	61	565	8.00	3.00	11.00	12.82	1.00	1.00
13	7	222.18	1	4.50	178.50	82	21.50	456.00	360.50	653	61	609	9.44	3.64	10.64	12.87	1.00	1.72
14	7	422.64	0	4.50	180.50	81	20.50	457.00	377.50	631	61	611	11.98	3.00	10.00	12.83	1.00	2.99
15	7	338.00	0	5.00	192.50	81	22.50	456.00	365.00	601	61	584	14.00	4.00	8.00	13.06	1.00	1.00
16	8	840.00	0	5.00	192.50	79	21.50	462.50	370.00	610	60	604	20.00	4.00	8.00	13.07	2.00	128.00
17	8	48.40	1	5.00	184.00	83	23.00	460.00	365.50	606	60	597	43.47	4.00	14.87	13.01	2.84	13.19
18	8	105.60	1	5.00	187.50	80	25.00	462.00	377.50	639	60	619	38.02	4.00	13.51	13.05	2.64	25.45
19	8	324.00	0	5.00	166.50	85	22.50	475.00	362.50	609	59	592	19.00	4.00	8.00	13.21	2.00	293.00

Table 2: On-board data and tribology data.

#	x	C_i	Air Fltr [kPa]	Boost Pres [kPa]	Eng Cool Temp (avg) [°C]	Eng Oil Fltr High & Cool>70 [kPa]	Eng Oil Pres Engine & Cool>74 [kPa]	Eng Oil Pres Engine & Cool>74 [kPa]	Lt Exh Temp [°C]	Rt Aftrclr Temp [°C]	Rt Exh Temp [°C]	Fe [ppm]	Si [ppm]	Na [ppm]	V100	Pb [ppm]	Cu [ppm]
1	75.04	0	4.50	191.00	82.00	23.00	429.00	348.00	625.00	61.00	626.00	1.03	0.21	9.00	13.32	1.00	1.00
1	76.26	0	5.00	188.00	86.00	24.00	419.50	338.50	636.00	62.00	638.00	1.18	0.24	9.00	13.32	1.00	1.00
1	77.49	0	5.00	185.50	83.00	23.00	422.00	362.50	620.00	61.00	609.00	1.32	0.26	9.00	13.32	1.00	1.00
1	78.71	0	5.00	176.00	86.00	24.00	421.50	340.00	650.00	62.00	593.00	1.47	0.29	9.00	13.32	1.00	1.00
1	79.93	0	5.00	180.00	84.00	24.50	428.00	352.00	637.00	62.00	594.00	1.62	0.32	9.00	13.31	1.00	1.00
1	81.15	0	5.00	185.00	85.00	21.50	426.00	346.50	642.00	62.00	606.00	1.76	0.35	9.00	13.31	1.00	1.00
1	82.37	0	5.00	184.00	85.00	22.00	425.50	337.00	649.00	62.00	599.00	1.91	0.38	9.00	13.31	1.00	1.00
1	83.59	0	5.00	175.00	75.00	24.50	433.00	353.50	618.00	61.00	616.00	2.06	0.41	9.00	13.31	1.00	1.00
1	84.81	0	5.00	179.50	76.00	21.00	442.50	372.00	626.00	63.00	624.00	2.21	0.44	9.00	13.31	1.00	1.00
1	86.03	0	5.00	178.50	84.00	21.00	423.50	352.00	653.00	65.00	624.00	2.35	0.47	9.00	13.31	1.00	1.00
1	87.25	0	5.00	180.50	82.00	21.00	423.00	367.00	666.00	64.00	627.00	2.50	0.50	9.00	13.31	1.00	1.00
1	88.47	0	5.00	159.00	81.00	20.50	431.50	370.50	589.00	61.00	584.00	2.65	0.53	9.00	13.30	1.00	1.00
1	89.69	0	5.00	178.50	85.00	21.50	424.00	350.00	666.00	63.00	611.00	2.79	0.56	9.00	13.30	1.00	1.00
1	90.91	0	5.00	128.50	83.00	21.00	430.00	356.50	600.00	60.00	536.00	2.94	0.59	9.00	13.30	1.00	1.00
1	92.13	0	5.00	180.50	82.00	22.00	431.50	365.00	622.00	61.00	590.00	3.09	0.62	9.00	13.30	1.00	1.00
1	93.35	0	5.00	188.50	84.00	21.50	421.00	354.50	645.00	61.00	593.00	3.24	0.65	9.00	13.30	1.00	1.00
1	94.57	0	5.00	194.50	83.00	24.00	434.50	346.50	605.00	61.00	603.00	3.38	0.68	9.00	13.30	1.00	1.00
1	95.79	0	5.00	179.50	85.00	23.50	418.00	347.50	643.00	61.00	604.00	3.53	0.71	9.00	13.29	1.00	1.00
1	97.01	0	5.00	179.00	85.00	23.50	419.50	351.00	629.00	61.00	594.00	3.68	0.74	9.00	13.29	1.00	1.00
1	98.24	0	5.00	191.00	78.00	24.00	432.50	359.00	611.00	61.00	597.00	3.82	0.76	9.00	13.29	1.00	1.00

Long Term Cost Optimization Decision Model

The following graphs form part of the decision model estimate results and are illustrated here for completeness sake.

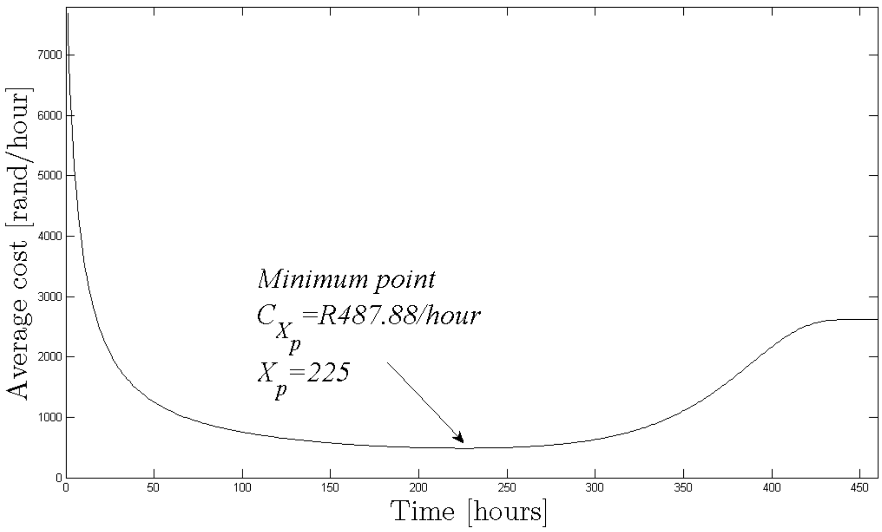

Figure 1: Cost function for the first inspection of the first event.

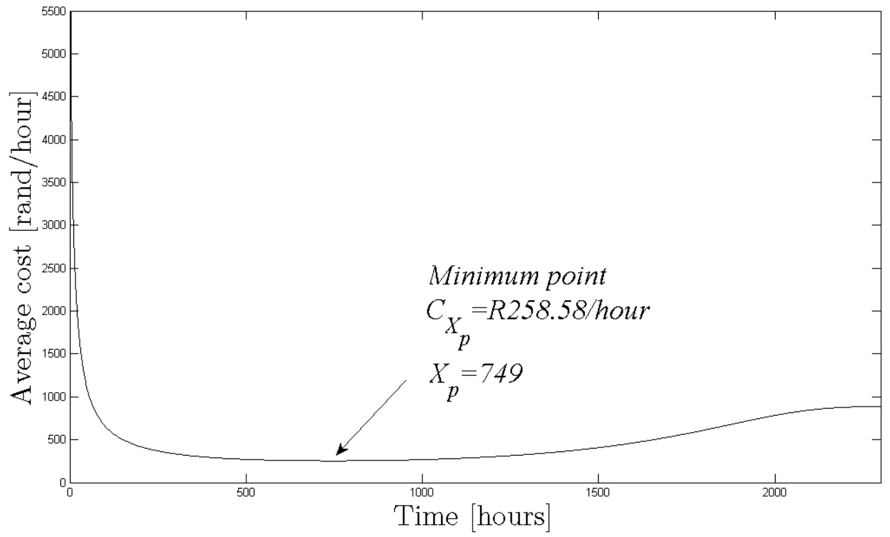

Figure 2: Cost function for the second inspection of the first event.

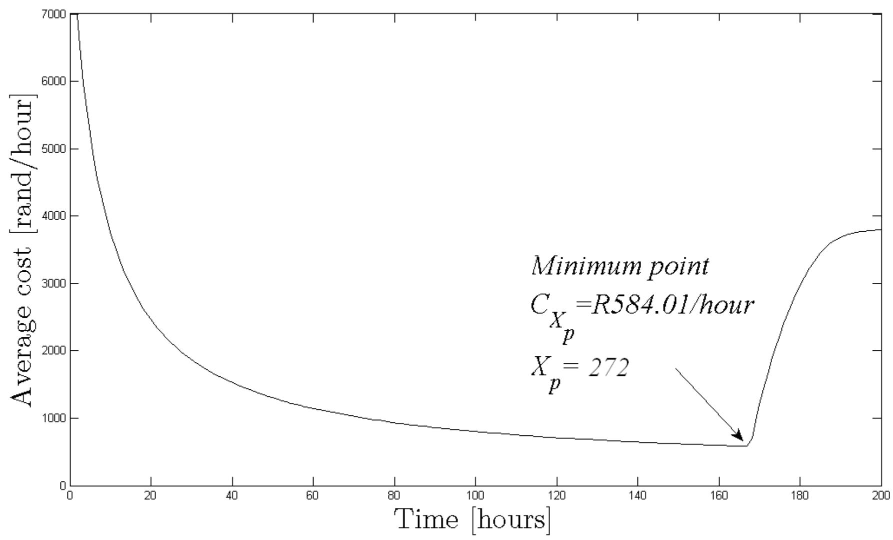

Figure 3: Cost function for the first inspection of the second event.

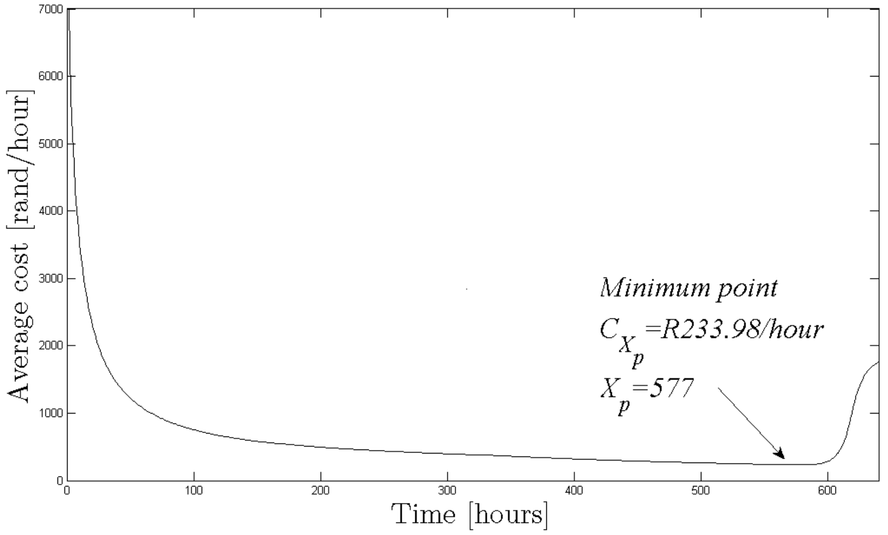

Figure 4: Cost function for the second inspection of the second event.

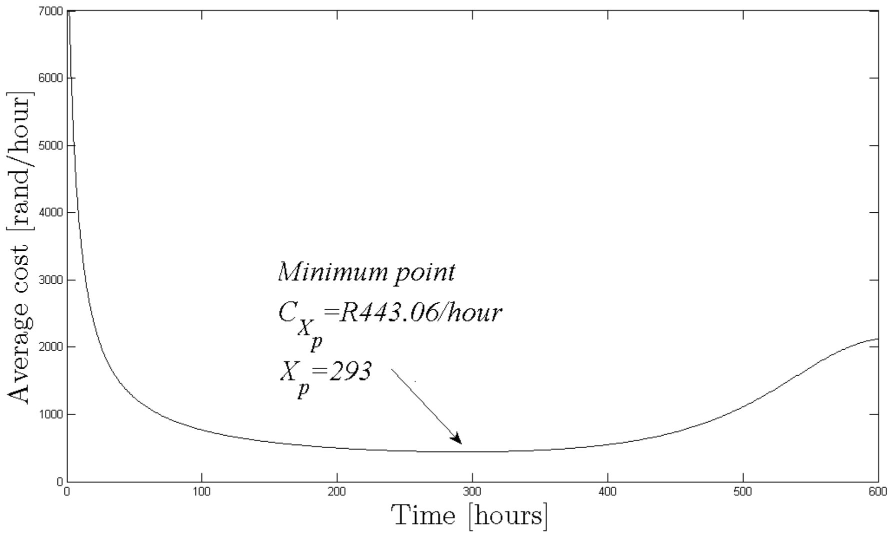

Figure 5: Cost function for the first inspection of the third event.

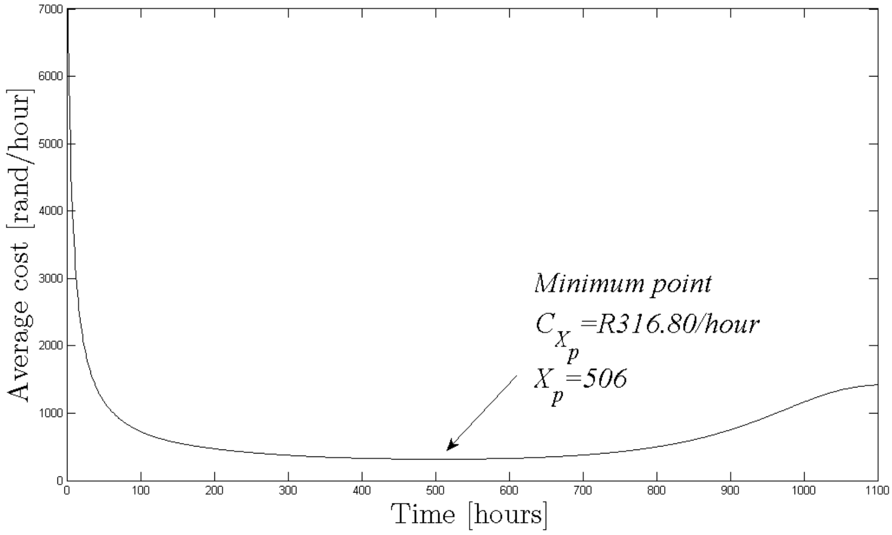

Figure 6: Cost function for the second inspection of the third event.

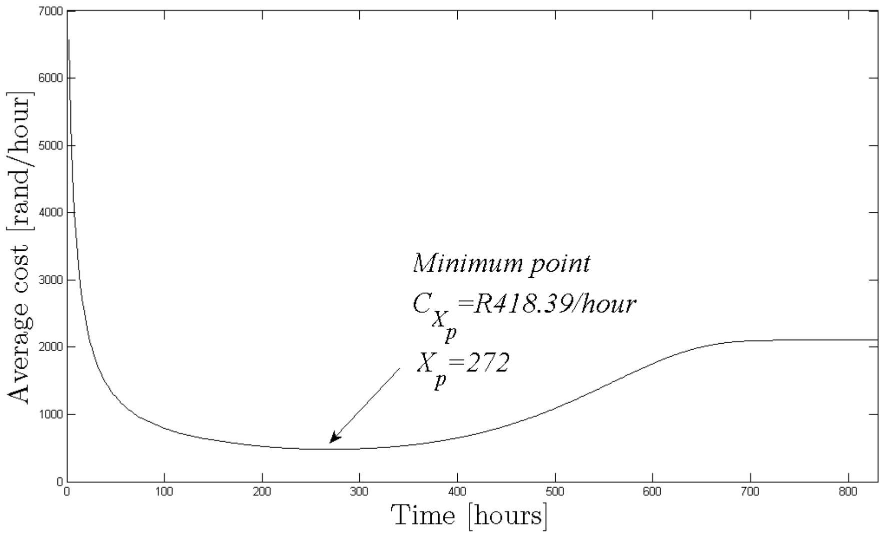

Figure 7: Cost function for the first inspection of the fourth event.

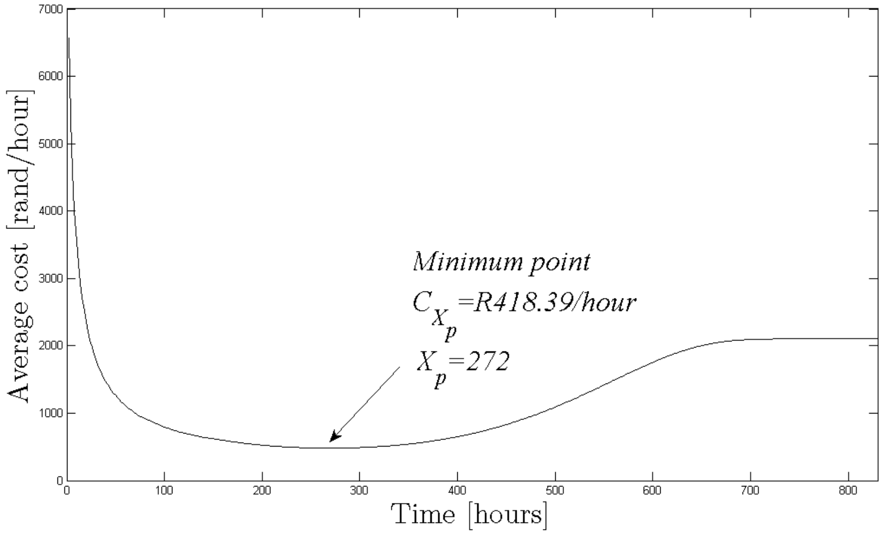

Figure 8: Cost function for the second inspection of the fourth event.

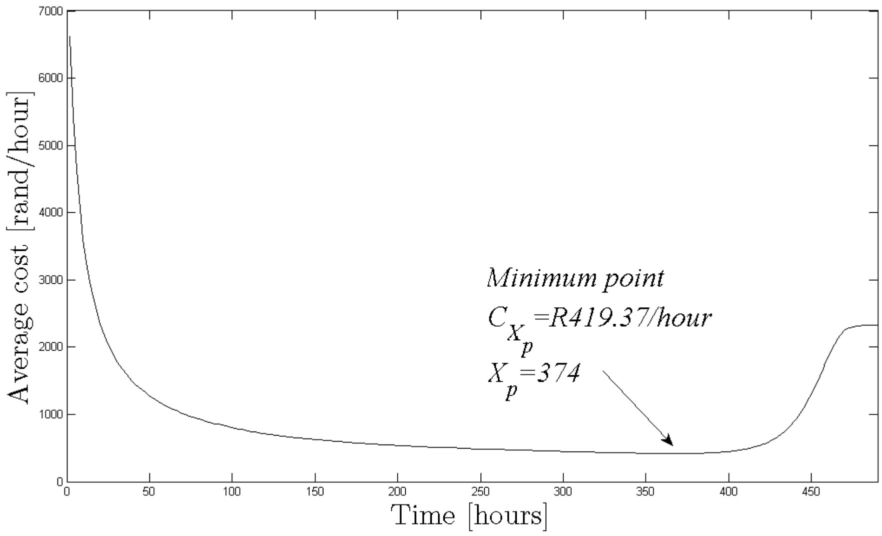

Figure 9: Cost function of the fifth event.