CRANFIELD UNIVERSITY

School of Aerospace, Transport and Manufacturing

IVHM Centre

PhD Thesis

Academic Year 2015 - 2016

Andrew John Wileman

An Investigation into the Prognosis of Electromagnetic Relays

Supervisor: Dr Suresh Perinpanayagam

June: 2016

Abstract

Electrical contacts provide a well-proven solution to switching various loads in a wide variety of applications, such as power distribution, control applications, automotive and telecommunications. However, electrical contacts are known for limited reliability due to degradation effects upon the switching contacts due to arcing and fretting. Essentially, the life of the device may be determined by the limited life of the contacts. Failure to trip, spurious tripping and contact welding can, in critical applications such as control systems for avionics and nuclear power application, cause significant costs due to downtime, as well as safety implications.

Prognostics provides a way to assess the remaining useful life (RUL) of a component based on its current state of health and its anticipated future usage and operating conditions. In this thesis, the effects of contact wear on a set of electromagnetic relays used in an avionic power controller is examined, and how contact resistance combined with a prognostic approach, can be used to ascertain the RUL of the device.

Two methodologies are presented, firstly a Physics based Model (PbM) of the degradation using the predicted material loss due to arc damage. Secondly a computationally efficient technique using posterior degradation data to form a state space model in real time via a Sliding Window Recursive Least Squares (SWRLS) algorithm.

Health monitoring using the presented techniques can provide knowledge of impending failure in high reliability applications where the risks associated with loss-of-functionality are too high to endure. The future states of the systems has been estimated based on a Particle and Kalman-filter projection of the models via a Bayesian framework. Performance of the prognostication health management algorithm during the contacts life has been quantified using performance evaluation metrics. Model predictions have been correlated with experimental data. Prognostic metrics including Prognostic Horizon (PH), alpha-Lamda (α-λ), and Relative Accuracy have been used to assess the performance of the damage proxies and a comparison of the two models made.

Keywords: *Electromagnetic Relays, Integrated Vehicle Health Management, Prognostics and Health Management, Condition Based Maintenance, Hybrid Prognostics, Physics-based Prognostics, Data-driven Prognostics, Kalman filtering, Particle Filtering.*

Acknowledgements

I would like to express sincere gratitude to my supervisor Dr Suresh Perinpanayagam for all his help and guidance whilst carrying out the work presented in this thesis. I would also like to extend my gratitude to my technical supervisor Dr James Whidbourne, for his input on mathematical modelling.

I would like to acknowledge my sponsors for this body of work, Meggitt PLC, Steve Parker and his colleagues and Thales (UK). As well constructive feedback provided by the staff and fellow students within Cranfield University IVHM Centre and its industrial partners during the Technical Reviews: Boeing, Rolls-Royce, BAE Systems, Alstom, and the Ministry of Defence.

Lastly, I would like to thank my parents for their support during this long journey.

Publications

Journal Papers

Wileman, Andrew J., and Suresh Perinpanayagam. "Integrated vehicle health management: An approach to dealing with lifetime prediction considerations on relays." *Microelectronics Reliability* 55.9 (2015): 2165-2171.

Wileman, A.J., Suresh Perinpanayagam and J.F. Whidborne. “A Sliding Window Recursive Least Square and Kalman-Filter based Model Prognostic for Assessing Relay Contact Life”, IEEE Transactions on Reliability, submitted for publication 18th February 2016.

Wileman, A. J., and S. Perinpanayagam. “A Regressive Model based Prognostic for Relay Contact Resistance Estimation”, Reliability Engineering & System Safety, submitted for publication 3rd March 2016.

Conference Proceedings

Wileman, A. J., and S. Perinpanayagam. "Failure mechanisms of radar and RF systems." *Procedia CIRP* 11 (2013): 56-61.

Wileman, A. J., and S. Perinpanayagam. "Prognostic Health Management of a RF Transceiver Chain." *Procedia CIRP* 11 (2013): 266-271.

Wileman, Andrew J., and Suresh Perinpanayagam. "A Prognostic Framework for Electromagnetic Relay Contacts." *The 2nd European Conference of the Prognostics and Health Management (PHM) Society, Nantes, France*. Vol. 5. No. 11. 2014.

Contents

List of Figures

List of Tables

ABBREVATIONS

AGO	Accumulated Generating Operation
ANFI	Adaptive Neuro Fuzzy Interface
ANNs	Artificial Neural Networks
AR	Auto-Regression
ARIMA	Auto-Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
CAA	Civil Aviation Authority
CBM	Condition Based Monitoring
CMMS	Computerised Maintenance Management System
CTW	Continuous Wave Transform
DD	Data Driven
DEC	Discontinuous Erosion Current
EA	Evolutionary Algorithms
EOF	End of Life
ES	Expert Systems
FADEC	Full Authority Digital Engine Controller
FBP	Fuzzy Back Propagation
FFT	Fast Fourier Transform
FL	Fuzzy Logic
FMEA	Failure Mode and Effect Analysis
FNN	Fuzzy Neural Networks
GM	Grey Model
HMM	Hidden Markov Model
HSMM	Hidden Semi-Markov Models
HUMS	Health and Usage Monitoring System
IAGO	Inverse Accumulated Generating Operation
ISHM	Integrated System Health Monitoring
IVHM	Integrated Vehicle Health Monitoring
KBM	Knowledge Based Models

KF	Kalman Filter
LCC	Life Cycle Cost
LSSVR	Least Square Support Vector Regression
MQE	Minimum Quantisation Error
MSE	Mean Squared Error
MTBF	Mean Time Between Failure
NFS	Neural Fuzzy Systems
PbM	Physics Based Model
PCA	Principle Component Analysis
PDF	Probability Density Function
PE	Persistently Exciting
PF	Particle Filter
PH	Prognostic Horizon
PHM	Prognostic Health Management
POF	Physics of Failure
RA	Relative Accuracy
RCM	Reliability Cantered Maintenance
RLS	Recursive Least Squares
RNN	Recursive Neural Network
ROI	Return on Investment
RUL	Remaining Useful Life
SIR	Sampling Importance Resampling
SPC	Statistical Process Control
SWRLS	Sliding Window Recursive Least Square
TF	Time Frequency
TRCU	Thrust Control Unit
TRL	Technology Readiness Level
UKF	Unscented Kalman Filter
α-λ	alpha lambda

Chapter 1 – Thesis Outline

Introduction

This chapter outlines the contents of the Thesis and why this work has come into fruition. This includes an overview of the background, aims and objectives of the study, as well as the contributions of the Thesis to new knowledge, the publications produced and finally an overview of the Thesis contents and layout.

1.1 Background and Motivation

In critical assets such as aerospace equipment, maintenance is a large expenditure for both the operator and the manufacturer, significant downtime causes a loss of business and ultimately money. As aircraft manufacturers move towards increasing aeroplane lifetimes (in excess of 30 years), it is important to increase the operational reliability of the avionic systems. Traditionally maintenance has occupied two strategies, namely corrective (sometimes called reactive or unplanned) and preventative maintenance (Williams, 2006), (Janasak and Beshears, 2007), (Byer, Hess and Fila, 2001) & (Hess, Calvello and Dabney, 2004).

Corrective maintenance is a remedial strategy and is carried out when a fault occurs and as far as possible is avoided within safety critical equipment such as avionics. Preventative maintenance is one such sub-group of proactive maintenance, where maintenance is performed on a periodical basis. Service periods are determined from the use of past historical data based on Mean Time Between Failures (MTBF) and not from any input from the machinery itself. The machinery is thus serviced at the scheduled period, whether it is necessary or not. Depending on the equipment, this can lead to improved system reliability, reduced cost of replacement, decreased system downtime and lead to a dedicated spares inventory. Generally, to make this viable, the components in question need to have an increasing failure rate throughout their lifetime, it should be noted that reliability is set at the design stage and degradation is a separate issue. The preventative cost needs to be less than the overall cost of a corrective action, at present this tends to suit high value assets, such as aerospace, defence etc.

Manufacturers and operators are constantly looking for a way to reduce this cost and achieve increased availability and downtime. Condition Based Monitoring (CBM) is a

maintenance program that recommends maintenance actions based on the information collected through condition monitoring of the component itself. CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviours of a physical asset. A correctly and effectively implemented CBM program, can significantly reduce maintenance cost by reducing the number of unnecessary scheduled preventive maintenance operations and has started to see usage on equipment such as aero-engines.

Future strategies to achieve an increased lifetime and maintainability of avionic systems will involve an intelligent prognostic capability implemented such that the ground crews are able to predict failures and the remaining service life of these systems and associated components. The aim is to build prognostics capability into avionic systems such that cost-effective, safe, reliable and robust systems can be built with a long service life.

CBM and prognostics form part of a methodology called Integrated Vehicle Health Management (IVHM) (Jennions, 2011), which allows unified capability of systems to assess the current or future state of system health, and integrate that system health within a framework of available resources and operational demands. IVHM is particularly attractive in high unit and high maintenance cost platforms such as aircraft and spacecraft and consists of following sub-categories:-

- Maintenance service offerings (e.g. CBM, Total Care, RCM)
- Business (e.g. Business models, IVHM mapping)
- System design
- Architecture
- Analytics (e.g. Diagnostics, Prognostics)
- Technologies (e.g. Structural Health Management (SHM))

IVHM enables many disciplines with an integrated framework.

1.2 Research Hypothesis

The main objectives of this thesis is to examine feasibility for the use of condition based monitoring (CBM) and prognostic implementation in high value safety critical assets within the aerospace and defence domain.

- The first objective in this thesis is therefore to examine the body of literature on CBM and prognostics already available and evaluate the various methodologies in terms of their applicability and accuracy for monitoring and prediction of component health and remaining useful life (RUL).

- The second objective will devise experimental testing to enable real world data to be collected and thus replicate the conditions experienced from the component in operation. This will result in any developed solutions having demonstrated capabilities and applicability to the conditions experienced in real world operation.

- The third objective will be to evaluate and develop prognostic algorithms for the data collected. This will involve an in depth study of the failure mode(s) in the monitored component and correlation between the data, algorithm and prediction.

- The fourth objective is the development of tailored, real time prognostic algorithms that are implementable at system level will be trialled and benchmarked to assess their performance.

The platform for the development of the work in this Thesis will be centred on the implementation of prognostics for relay outputs on the Full Authority Digital Controller (FADEC) for a range of aircraft currently being developed. Within safety critical systems, relays are still favoured over their more reliable solid state counterparts due to their ability to completely isolate the circuitry they are switching, with no leakage current; hence their continued usage within equipment that is deemed safety critical, such as aerospace, nuclear and rail signalling.

The FADEC provides control and optimisation of the aircrafts engine performance by using a digital computer to analyse a multitude of operating parameters and cannot be manually overridden. The unit shown in figure 1.1 is airframe-mounted, performing pre-defined functions; including signal processing, communication, logic operation and built-in-test (BIT) and provides a dedicated interface between the engine and the aircraft. Being a redundant system, the board is replicated on two channels A and B.

The research carried out in this thesis is to investigate the viability of a prognostic implementation up to Technology Readiness Level 4 (TRL4), (NASA, 2015) within the FADEC. The prognostic algorithm must be capable of being utilised within current embedded technology and all non-board sensors are to be non-intrusive to satisfy the CAA (Civil Aviation body).

It is envisaged at this stage that the prognostic capability will focus on the area of monitoring the life of the relays providing the control output to the Full Authority Digital Engine (or electronics) Control (FADEC) and the Thrust Control Unit (TRCU).

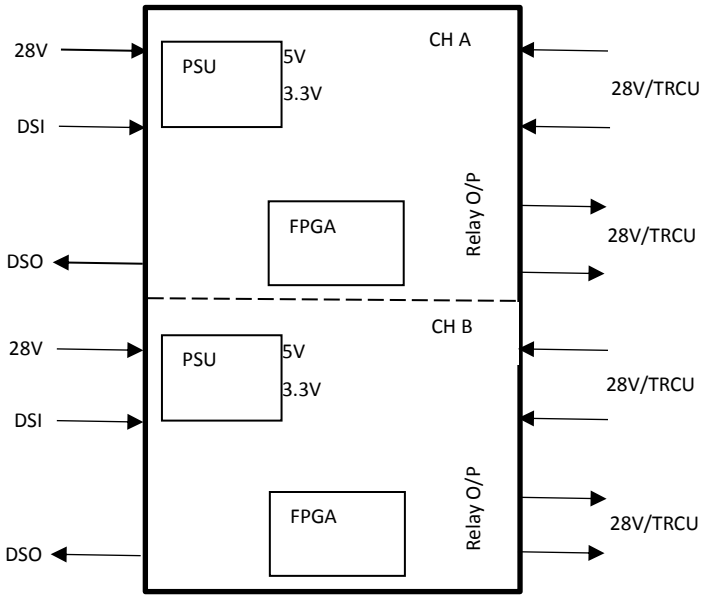


Figure 1.1. Overview of the FADEC layout

1.3 Contributions of this Thesis

The main contributions of this Thesis are outlined below and in the conclusion of each of the relevant chapters.

- A method has been developed for the obtainment of real time measurements to assess the life of the relay using parameters such as contact resistance.
- A data set has been produced that describes relay failure. This has been broken down into sections to enable the construction of a visual analytics model to enable to application of prognostic methods.
- The failure modes found within the relay have been structured into a flowchart and have been pictured as an early decision tool.

- The development of a model for the estimation of secondary parameters within the relay such as coil current, contact bounce and contact position.
- A model to enable the primary effects from arcing damage to be examined. The model incorporates the ability to look at heat flow through the contacts, electrical voltage across the contact due to heating and mass loss from the contact.
- The mass loss estimation which contributes towards the development of a state space physics of failure model.
- Estimation of RUL from this model.
- The development and assessment of a real time Grey Model prediction algorithm for prognostics.
- A novel and computationally efficient technique has been developed for real time monitoring based upon modelling of degradation using system identification techniques.

1.4 Thesis Layout

The organisation of this Thesis is as follows:

Chapter 2 examines the primary failure modes in relays and then goes on to look at the mechanisms that cause the failure and useful parameters that may be used to develop prognostics.

Chapter 3 provides a detailed literature review of maintenance, condition monitoring and prognostic strategies. The various prognostics approaches are examined in detail and initial conclusions are drawn on the suitability of methodologies for use in the application of relays.

Chapter 4 focusses on the methodology used for the data extraction. The consideration of a suitable metric is first explored which reflects accurately the degradation of the relay contacts. This is followed by a discussion of the experimental apparatus devised and constructed within the laboratory to collect the data. Analysis of the measurement and process noise within the data follows. Lastly, the raw data is discussed in terms of patterns and features that may be applicable for prognostics as well as filtering methodologies to assist with trend clarification.

Chapter 5 develops a series of models for the development of a physics based model. Firstly, a model of the secondary characteristics such as coil current, contact bounce

and contact position is developed. Secondly, an arcing model is developed and a numerical solution of the heat equation is used to explore temperature distribution and mass loss from the contact. Finally in this chapter, a simple state space model is suggested and used to predict the RUL.

Chapter 6 looks at using real time data driven methods and proposes a new approach for devising the prognostic model in real time, using a sliding window and modelling methodology based upon system identification.

Chapter 7 presents the results from the models derived in chapters 5 and 6 respectively. The Physics Based Model is firstly analysed and discussed and then compared to the Data Driven model. This chapter starts by outlining current prognostic metrics and the results are evaluated according to these metrics. The chapter concludes by discussing the outcomes of the results.

Chapter 8 concludes the Thesis by reviewing the results of the previous chapter and suggests future work based upon the outcomes.

Chapter 2 - Relay Failure Modes and Parameters

Introduction

In order to enable the development of a prognostic methodology for electromagnetic relays, a thorough understanding of the way the device degrades is essential. This chapter firstly examines the primary failure modes in relays. It then goes on to look at the mechanisms that cause the failure, also useful parameters that may be used to develop prognostics.

The electromagnetic relay has been around for a very long time, approximately 160 years and is essentially an electrically operated switch. The basic principle of most relays is to use an electromagnet to operate a mechanical switching mechanism; solid state versions of the device are also available, but will not be considered in any depth here. Relays are used for the control of circuits via a low power signal and offer complete isolation between the control and the controlled circuit. The other advantages are their ability to deal with high surge currents and high voltage spikes as well as having no leakage current. However, their main disadvantage is the life expectancy, which is low, compared with their solid state counterpart etc.

Relays have many applications, amongst the first uses were in telephony and telephone exchanges, as well as early computing. Modern uses are still many and varied, with applications such as amplifying digital signals, switching a large amount of power with a small operating power; industrial control of machine tools, transfer machines, and other sequential control; detection and isolation of faults on transmission and distribution lines by opening and closing circuit breakers (protection relays); isolation of the control circuit from the controlled circuit and logic functions.

Within safety critical systems, relays are still favoured over their more reliable solid state counterparts due to their ability to completely isolate the circuitry they are switching, with no leakage current; hence their continued usage within equipment that is deemed safety critical, such as aerospace, nuclear and rail signalling. However, relays are known for limited reliability due to the mechanical wear of internal switching elements, and essentially the life of the relay, may be determined by the life of the contacts. Failure to trip, spurious tripping and contact welding, can in safety critical applications such as control systems for avionics and nuclear power stations, as well as signaling in rail networks, cause significant costs due to downtime and safety implications.

The use of prognostics provides a way to assess the Remaining Useful Life (RUL) of an electromagnetic relay based on its current state of health and its anticipated future usage and operating conditions. However, the development of a methodology to enable this to be implemented is complex, due the nature of devices failure mode. In order to progress, the physics of degradation needs to be understood.

2.1 Relay Design and Operation

The design of electromechanical power relays is basically the same regardless of whether it is a miniature PCB relay or an industrial power relay. The most important components are; the contacts, magnetic system and casing as pictured in figure 2.1 and described below.

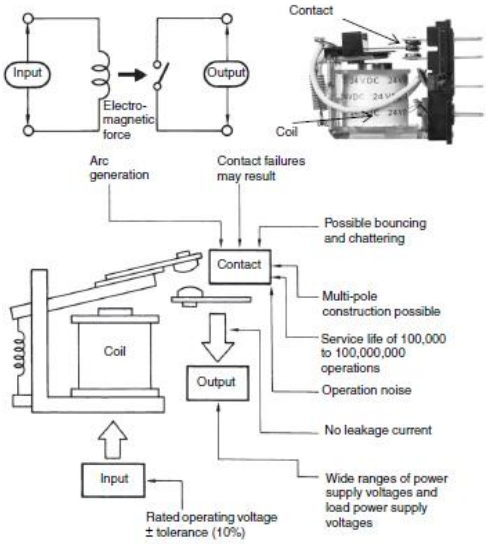


Figure 2.1. Basic components of an electromagnetic relay (image from Omron Components, 1990)

The contacts may be fixed or moving (contacts being moved by the magnetic system (motor) to switch the load circuit), and are usually mounted on contact springs (which hold the contacts but are sufficiently flexible to allow the contacts to move).

A magnetic system which consists of; a coil (to generate the necessary magnetic field to actuate the armature and the contacts), a core, a yoke (to establish the magnetic circuit), an armature (the moving part of the magnetic system which closes and opens the magnetic circuit and acts via a comb or actuator on the moving relay contacts), and lastly the return spring (to establish a defined position of the magnetic and contact system in case the coil is not energized).

Finally, the mechanical components, including a case, a base (to protect the relay against external influences and for protection against electric shock), insulation (within

the relay to separate the primary circuit from the secondary side and to provide the required insulation), an actuator (used in some relay designs to translate the motion of the magnetic system to the contact system (moving contacts)), pins or terminals (to connect between the contact system and the load) and lastly, a mounting device.

2.2 Failure and Degradation Modes in Relays

At a first glance, the relatively high component count may imply a large scope for failure, but years of development and design has ensured the overall reliability is high. Data sheets will allude to power and general relays having an electrical (contact) life expectancy usually in excess of 100,000 operational cycles minimum with a resistive loading, this decreases significantly if the loading is inductive. This may be compared to the mechanical life expectancy, which is in the order of one million and in some cases 10 and 100 million operations.

The reason the electrical life is so low compared with the mechanical life is because the contact life is application dependent. The electrical rating applies to contacts switching at their rated loads (Slade, 2013). If however, a set of contacts is used to switch a load less than the rated value, the contact life again, may be significantly higher. The rated electrical life also takes into account arc destruction of the contacts. Arc suppression may be used to lengthen the life of the contact.

The attribution of failure in relays can be divided into two main areas, firstly from contamination and secondly the wear of internal switching elements. Failure due to contamination is due to numerous causes, but again may be divided into two main camps: metallic and non-metallic. Metallic contamination constitutes to unwanted foreign particles that may build up within the device during use. It also takes into account chemicals that may be encountered during manufacture, all of which may impair normal mechanical device operation. Secondly, contamination may occur from non-metallic or gaseous deposits presenting as a film which creates open circuits when it periodically deposits itself on the contacts.

The second major and more detrimental cause of premature early life relay failure is the electrical wear of internal switching elements. In fact, it may be said ‘the life of a relay is essentially determined by the life of its contacts.’

High-sustained currents, high in-rush currents and voltage spikes lead to the degradation of the contacts. The high currents and voltages sources tend to be load

dependent, with inductive loads creating the highest voltage and current spikes due to the low starting resistance compared to operating resistance. This is especially true for lamp filaments and motors, which is why de-rating is more severe for these types of loads. Degradation may be further accelerated if contamination or pitting is present on the contact.

Figure 2.2. The effects of contact degradation

Other elements within the relay may also suffer from physical wear. Springs, which are often present to provide mechanical return when a switching current is not applied, may lose their resilience over time. The alignment of the contact can also be a cause of premature failure, as well is if the coil beaks down or goes open circuit.
A summary of failure is presented below (Fujitsu Components, 2009).

Parts	**Stress**			**Failure symptoms**	**Failure Mode**
	Status	**Environment**	**Duration**		
Contact	Voltage, Current, Surge voltage,	Temperature, External Vibration, Humidity, External Shock, Dust, Gas	Long duration without breaking, Intermittent	Transfer and wear of contact material due to arc discharge, Weld and bridging of contact Sticking of contact Corrosion (oxidation, sulfurization etc.), Foreign matter, Deposits	Poor release Poor contact Increased contact resistance Noise Change in operate/release time Poor dielectric strength
Winding	As above	As above	As above	Corrosion Galvanic corrosion Foreign matter (dust etc.) Voltage fluctuation Vibration of lead wire	Breakage of coil Burning of coil Poor working release operation Change in operate/release time Change in operate/release voltage Malfunction
Structural parts	As above	As above Excessive external shock	As above	Slip-off and wear of contact piece, Fatigue and creep of spring, Abnormal wear and loosening, Seizure.	Poor contact Poor release Change in operate/release time

				Deterioration of organic material, Deposition of worn contact material powders, Corrosion and galvanic corrosion, Foreign matter (dust etc.)	Change in operate/release voltage Insulation resistance
Enclosure	As above, Chemicals	As above, chemicals	As above	Damage by external force Change in chemical properties	Damage (cracks etc.)

Table 2.1. A summary of failure symptoms and modes (Fujitsu Components, 2009).

From the above discussion, one may conclude that the development of a prognostic model should be focused on the electrical life of the contacts, due to the life expectancy being in the order of magnitude of ten times less than the mechanical counterparts. This should not however exclude the effects of wear of the other components, in fact, parameters such as contact force, operation/release times and contact resistance will depend upon windings, structural parts, etc.

2.3 Electrical Life of Contacts

A look at a solid surface under a microscope will show that even the smoothest appearing surfaces are in fact rough. The micro-surface will be composed of peaks and valleys, whose height variations, shape and other geometric considerations vary considerably.

When a contact is made between two metals, surface asperities of the contacting members will penetrate the natural oxide and other surface contaminant films, establishing localized metallic contacts and, thus, conducting paths. As the force increases, the number and the area of these small metal to metal contact spots will increase as a result of the rupturing of the oxide film and extrusion of metal through the ruptures. These spots, termed a-spots, are small cold welds providing the only conducting paths for the transfer of electrical current. A direct consequence of this is a porous contact where infiltrating oxygen and other corrosive gases can enter to react with the exposed metal and reduce the metallic contact areas. This will eventually lead to disappearance of the electrical contact, although the mechanical contact between the oxidized surfaces may still be preserved (Braunovic et al., 2006).

Current flowing in an electrical contact is funneled and channeled through the separate a-spots (see figure 2.3 below). This constriction of the electric current by the a-spot

formation reduces the overall volume of material used for electrical conduction and this leads to increase in the electrical resistance. The increase in resistance is defined as the constriction resistance of the contact interface.

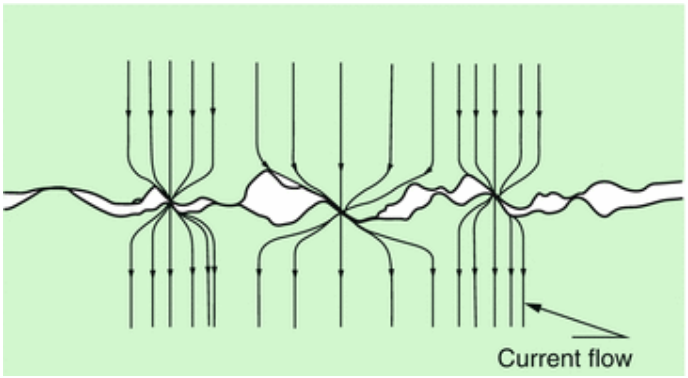


Figure 2.3. Illustrating the effect of constriction resistance on current flow

2.4 Mechanics of a-spot Formation

An electrical contact has a real contact area made up from the a-spots as opposed to the apparent contact surface which is made up of surface asperities and dirt (Holm, 2000). A generalization to make calculations simpler is to assume that a-spots are circular in nature. The simplification is illustrated below in figure 2.4 below.

Several references (Holm, 2000 & Smythe, 1968) treat the mathematical analysis of circular a-spot and the constriction of current flow and the results are summarized.

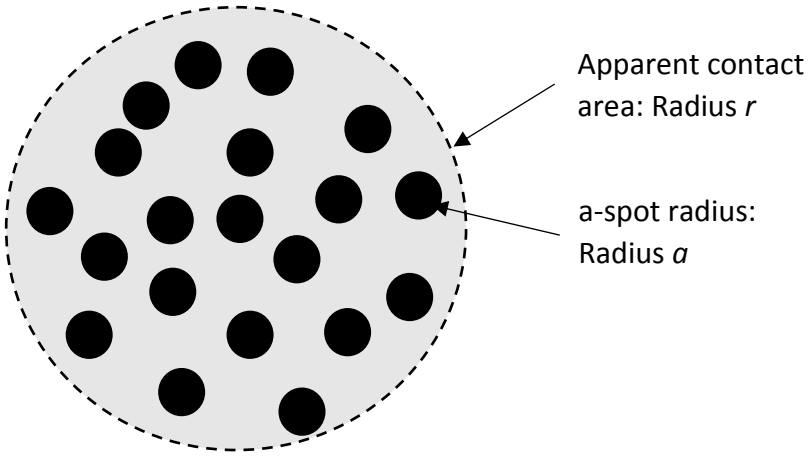


Figure 2.4. Representing the simplified depiction of a-spots and how the apparent contact radius is reduced

The contact resistance R_c of an a-Spot is represented as the sum of the constriction resistance R_s and film resistance by using the following equation reported by (Holm, 2000).

$$R_c = \frac{\rho}{2a} + \frac{\rho_f d}{\pi a^2} \quad (1)$$

where ρ indicates the resistivity of a contact material, *a* indicates a contact radius, ρ_f indicates the resistivity of a film, and *d* is the thickness of the film. It is reported that the constriction resistance in Ohms is dominant if the load is over 10 N (Saitoh et al. 2007).

An approximation to the above equation was also proposed by Holm for the contact resistance illustrated in the above figure 2.4.

$$R_c = \frac{\rho}{2r} + \frac{\rho}{2na} \quad (2)$$

where *r* indicates the radius of an apparent contact area, *n* indicates the number of a-spots, and *a* indicates the radius of a-Spot.

Greenwood released the following contact-resistance approximation for multiple contacts by considering the interaction between a-Spots (Greenwood 1966).

$$R_c = \frac{\rho}{2\sum a_i} + \frac{\rho}{\pi}\left(\sum\sum_{i\neq j} \frac{a_i a_j}{S_{ij}}\right)\frac{1}{(\sum a_i)^2} \quad (3)$$

where S_{ij} indicates the distance between a-spots, and a_i, a_j indicates the radius of a-spot.

The difference between the approximation in equation (2) and Greenwood's equation (3) for the contact resistance as a function of the changing radius of the a-spots is illustrated below.

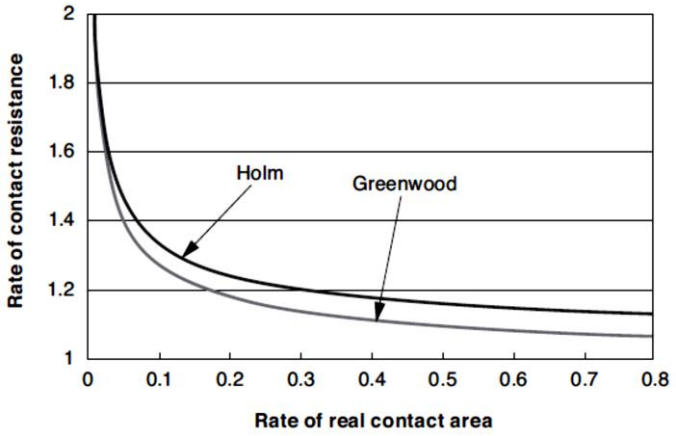


Figure 2.5. Relationship between rate of contact resistance and rate of real contact area in apparent contact area.

The horizontal axis of figure 2.5 indicates the rate of the real contact area in the apparent contact area, and the vertical axis indicates the rate of the contact resistance compared with that the rate of real contact area is 1 Ω. When the rate of the real contact area is above 0.2, the rate of contact resistance increases only about 1.14-

1.25 times. It may be assumed that the constriction resistance of apparent contact area is almost same as that of the real contact area presented by equation (2) and hence the following equation is a good approximation.

$$R_c = \frac{\rho}{2a} \quad (4)$$

The above equations are important in the design of electrical contacts and are true for monometallic contacts even when there are numerous clusters of a-spots that are not necessarily circular.

The contact resistance R_c between two conductors of resistivity ρ_1 and ρ_2, held together with a force F, is given as (Holm, 2000) as

$$R_c = \frac{\rho_1+\rho_2}{4}\sqrt{\frac{\pi H}{F}} \quad (5)$$

where H again, is the Vickers' micro-hardness of the softer of the two materials and F is the contact force. Again, because the metals are not clean, the passage of electric current may be affected by thin oxide, sulphide, and other non-organic films usually present on metal surfaces. Consequently, the total contact resistance of a joint is a sum of the constriction resistance (R_s) and the resistance of the film (R_f).

The contact resistance is the most important and universal characteristic of all electrical contacts and is always taken into account as an integral part of the overall circuit resistance of a device. Therefore, although it is significantly smaller as compared with the overall circuit resistance, the changes in the contact resistance can cause significant malfunctions of the device. This is because the contact resistance can vary significantly with the changes in the real contact area, contact pressure variations, resistive film non-uniformity, and other factors (Braunovic et al., 2006).

2.4.1 Temperature of an Electrically Heated a-spot

If the electrical current through the constriction of the a-spot is sufficiently small and the effect on contact resistance due to the heat generated is negligible, the definition of the constriction resistance presented in equation (4) is valid. Observing these conditions, the voltage drop across the constriction produced by an electric current I is

$$V = IR_c = I\rho/2a \quad (6)$$

However, a thermal gradient is produced normal to the constriction when significant Joule heating is experienced through the a-spot due the crowding of current lines. The gradient complicates the relationship between the dimension of the a-spot, electric current and the voltage drop across the contact.

On passing an electric current, the temperature of an a-spot rises very quickly to near equilibrium temperature, this is given by a thermal transient time. In slow moving and stationary contacts, the equilibrium temperature is generally only of interest. Thermal transients are only taken into account in rapidly moving contact applications such as brush contacts, high-power and high-frequency electrical connections (Braunovic et al., 2006).

2.4.2 Voltage-Temperature Relationship

Assuming that the outer surfaces of the conductors are thermally insulated from the external environment, the heat produced within an a-spot can, thus, be dissipated only by conduction through the bodies in contact. The electric and thermal current lines follow the same path under these conditions and hence the electric potential and isothermal surfaces within the conductor coincide (Kohlrausch, 1900), (Diesselhorst, 1900), (Greenwood and Williamson, 1958).

This unique relationship between electric and thermal current leads to the simple relationship between the voltage drop V across the contact and the maximum temperature T_m in the contact interface as

$$V = \left\{2 \int_{T_1}^{T_m} \lambda_1 \rho_1 dT\right\}^{\frac{1}{2}} + \left\{2 \int_{T_2}^{T_m} \lambda_2 \rho_2 dT\right\}^{\frac{1}{2}} \quad (7)$$

where λ and ρ are respectively the thermal conductivity and electrical resistivity of the conductors, the subscript numbers refer to the two conductors in contact and T_1 and T_2 refer to the bulk temperatures of the contacting bodies (Slade, 2013). The quantities λ and ρ generally vary with temperature. Due to the electric current flow lines being constricted the most within the confines of the a-spot and its immediate vicinity, the maximum temperature T_m occurs here. For the case of monometallic contact where $\lambda_1 = \lambda_2 = \lambda$ and $\rho_1 = \rho_2 = \rho$, T_m occurs precisely at the a-spot and the relationship in (7) reduces to

$$V = 2\left\{2 \int_{T_1}^{T_m} \lambda \rho dT\right\}^{\frac{1}{2}} \quad (8)$$

In the temperature range where λ and ρ vary little with temperature, then the above equation reduces to a simpler form giving a relationship between the voltage and temperature (V-T) relationship for monometallic electrical contacts.

$$T_m - T_1 = \frac{V^2}{8\lambda\rho} \quad (9)$$

The left hand side is the contact super-temperature, which reflects the deviation of the a-spot temperature T_m from the bulk connector temperature T_1. This equation loses its validity when the super-temperature is large, in excess of several tens of degrees as it is based on the consistency of the electrical resistivity and thermal conductivity over a small range. One important thing to observe about this equation is that it is independent of dimensional constraints such as size and shape and purely uses the material properties λ and ρ making it suitable for use with any thermal insulated body capable of passing an electrical current.

As the above equation is only valid over a small temperature range, equation (7) can be expanded to incorporate the dependence of the thermal conductivity and electrical resistivity on temperature over a greater range by introducing the relationships $\lambda = \lambda_0(1 - \beta T)$ and $\rho = \rho_0(1 + \alpha T)$ (Holm, 2000), (Greenwood and Williamson, 1958). The thermal conductivity of many metals usually decreases as the temperature increases, by using the values for β and α which are readily tabulated for many metals at 20 °C, equation (7) becomes (Slade, 2013).

$$V^2 = 8\lambda_0\rho_0\left(\{T_m - T_1\} + \frac{\{\alpha-\beta\}}{2}\{T_m^2 - T_1^2\} - \frac{\alpha\beta}{3}\{T_m^3 - T_1^3\}\right) \quad (10)$$

The above equation gives a method of calculating the voltage drop across the contact given the contact deviation due to heating. In all cases the contact temperture deviates significantly from the bulk temperature only when the voltage drop across the contact exceeds aproximately 10 mV. A potential drop exceeding in excess of 10 mV can lead to the production of temperatures that easily lead to softening and melting of the contact material. Table 2.2 below, gives the voltage temperature relationship for common electrical contact materials (Holm, 2000). The limitations of this equation were examined by (Timsit, 1983) experimentally and shown to be valid as long as the average diameter of the a-spot is larger than the mean free path of free electrons in the contacting bodies.

Material	Heat Capacity ($Jm^{-3}C^{-1}x10^6$)	V_s Softening (V)	V_m Melting (V)
Al	2.4	0.1	0.3
Fe	3.6	0.19	0.19
Ni	3.9	0.16	0.16
Cu	3.4	0.12	0.43
Zn		0.1	0.17
Mo	2.6 at 20°C	0.25	0.75
	3.4 at 1500°C		
Ag	2.5	0.09	0.37
Cd	2		
Sn	1.65	0.07	0.13
Au	2.5	0.08	0.43
W	2.7 at 20°C	0.4	1.1
	3.5 at 1400°C		
	3.9 at 2100°C		
Pt	2.8	0.25	0.65
Pd	2.7		0.57
Pb		0.12	0.19
60Cu, 40Zn	3.2		0.2
60Cu, 40Sn	3		0.15
Stainless Steel	3.9	0.27	0.55
WC	3	0.6	

Table 2.2 Voltage for softening (V_s) and Melting (V_m) of Common Electrical Contact Materials (Holm, 2000)

2.4.3 The Wiedemann-Franz Law

Equation (10) above, although an accurate representation of the V-T relationship over which the resistivity and thermal conductivity varies linearly with temperature is somewhat cumbersome to use. The Wiedeman-Franz law states that the variations with temperature of the thermal conductivity and electrical resistivity of metals are such that λ and ρ are related by the expression

$$\lambda\rho = LT \tag{11}$$

where L is the Lorentz constant (2.45 x 10^{-8} V^2K^{-2}) and T is the absolute temperature. The above equation is valid if thermal conduction and electrical resistivity arise from electronic transport in the metals. If the relationship in (9) holds, then equation (8) gives the V-T relation as

$$V^2 = 4L(T_m^2 - T_1^2) \tag{12}$$

which is independent of the material in the contact. The super-temperature $(T_m - T_1)$ in equation (12) however cannot be evaluated from the voltage drop. Therefore, by redefinition and rearrangement of the terms in equation (12), the equation may be reduced to the form of equation (9). Defining $\frac{(T_m - T_1)}{2}$ as the average temperature T in the contact, equation (12) can be written as

$$\lambda\rho = \frac{L\{T_m + T_1\}}{2} \tag{13}$$

It follows that $\{T_m^2 - T_1^2\} = \{T_m - T_1\} \times \{T_m + T_1\} 2\lambda_{avg}\rho_{avg}/L$. Substituting into the equation (9) gives

$$T_m - T_1 = \frac{V^2}{8\lambda_{avg}\rho_{avg}} \tag{14}$$

now with the V-T relation recovered in terms of the physical parameters λ and ρ evaluated at the average connector temperature (Slade, 2013).

2.4.4 Temperature Distribution in the Vicinity of an a-spot

The Wiedeman-Franz law does not necessarily apply that the maximum contact temperature occurs at the location of the physical interface.

The calculation of the temperature distribution within an a-spot depends upon the geometrical coincidence of the equipotential and isothermal surfaces within the volumes of the conductors. Work carried out by (Holm, 2000) and (Greenwood and Williamson, 1958) found that the maximum contact temperature T_m and the current *I* are related by the expression

$$R_c I = \rho_{0,1} \int_{T_1}^{T_m} \left\{ 2 \int_{T}^{T_m} \lambda_1 \rho_1 \, dT \right\}^{-\frac{1}{2}} + \lambda_1 \, dT + \rho_{0,2} \int_{T_2}^{T_m} \left\{ 2 \int_{T}^{T_m} \lambda_2 \rho_2 \, dT \right\}^{-\frac{1}{2}} \lambda_2 dT \tag{15}$$

where T_1 and T_2 are the bulk temperatures of the materials 1 and 2 respectively, R_c is the 'cold' contact resistance, $\rho_{1,0}$and $\rho_{2,0}$ are the cold resistivity's of materials 1 and 2 defined through the dependence on temperature of electrical resistivity, e.g. $\rho_0(1 + \alpha T)$, and $\lambda_1$and λ_2 are the thermal conductivities of the materials respectively. Equation 12 shows the maximum temperature developed in a contact is related to the voltage drop developed by the same current, if the identical contact had remained cold, since $R_c I$ is the 'cold' voltage drop.

Equation (15) can be approximated by again using the temperature dependence $\lambda = \lambda_0(1-\beta T)$, and if β is sufficiently smaller than α (e.g. β~α/20 for copper), the integral in (14) is given approximately by

$$\int_{T_1}^{T_m}\left\{2\int_{T}^{T_m}\lambda\rho dT\right\}^{-\frac{1}{2}}\lambda dT' = \left(\frac{\lambda_0}{\{\alpha-\beta\}\rho_0}\right)^{\frac{1}{2}}\left\{\frac{\alpha}{\alpha-\beta}\right\}cos^{-1}\left(\frac{1+\{\alpha-\beta\}T_1}{1+\{\alpha-\beta\}T_m}\right) \quad (16)$$

For a monometallic junction and bulk temperatures T_1 and T_2 of 0 ℃, R_c is given as $\rho_0/2a$, equation (16) now becomes

$$I = 4a\left(\frac{\lambda_0}{\{\alpha-\beta\}\rho_0}\right)^{\frac{1}{2}}\left\{\frac{\alpha}{\alpha-\beta}\right\}cos^{-1}\left(\frac{1}{1+\{\alpha-\beta\}T_m}\right) \quad (17)$$

An equipotential surface at temperature T and located at a fixed axial distance μ from the constriction, R_C can be given as $(\rho_0/\pi a)tan^{-1}(\mu/a)$, then equation (17) becomes (Slade, 2013) & (Greenwood and Williamson, 1958).

$$I = \frac{2\pi a}{tan^{-1}\left\{\frac{\mu}{a}\right\}}\left(\frac{\lambda_0}{\{\alpha-\beta\}\rho_0}\right)^{\frac{1}{2}}\left\{\frac{\alpha}{\alpha-\beta}\right\}cos^{-1}\left(\frac{1+\{\alpha-\beta\}T}{1+\{\alpha-\beta\}T_m}\right) \quad (18)$$

Equating equations 17 and 18 gives the dependence of the temperature T at the isothermal surface located at a distance μ from the constriction and is given by

$$cos^{-1}\left(\frac{1+\{\alpha-\beta\}T}{1+\{\alpha-\beta\}T_m}\right) = \frac{2}{\pi}tan^{-1}\left\{\frac{\mu}{a}\right\}cos^{-1}\left(\frac{1}{1+\{\alpha-\beta\}T_m}\right) \quad (19)$$

2.5 Electrical Arcing

An arc is produced from stored energy in a circuit due to the inductance L. If the current was suddenly to drop to zero in a circuit by the parting of the electrical contacts, then the stored energy in circuit inductance would result in large over voltages given by

$$V = -L\frac{dI}{dt} \quad (20)$$

In a DC circuit the duration of the arcing time is related to the magnitude of the arc voltage U_A compared with the circuit voltage U_C. When $U_A > U_C$ a finite time is required to dissipate the $\frac{1}{2}LI^2$ energy stored in the circuit inductance.

Electric arcs, although destructive, are necessary; they allow the smooth transition of current in a circuit to go steadily to zero. The designers of switching equipment have to incorporate into their design measures to handle these high voltages, the presence of arcing however, for the most part limits the values of these back emf's to a maximum of two times the circuit voltage. As well as the control of the dissipation of energy,

arcing can be useful in some contact designs to clear organic matter and corrosive oxides from the mating faces.

To allow the development of a state model for contact degradation, a clear understanding of arc formation is needed. The study of arcing in contacts from literature will be beneficial in forming an overall model that relates the excess circuit energy into arcing, into heat and finally into the degradation of the contact overall.

Different materials will have different arc voltages and currents, the arc will ignite if both the minimum arc voltage and current are exceeded. Table 2.3 shows these voltages and currents for several different materials (Slade, 2013). The melt voltage is also shown for the material due to arcing. Arcing should, as much as possible, be suppressed due to the damage it causes, however, it can be necessary to allow some arcing as a method to remove sulfidation, oxidation and contaminates from the contact surface in some materials.

As well as arcing, sparking may cause damage at voltages and currents less than those required for arc ignition. The spark is due to capacitive discharge, and compared to an arc is weak, and contributes less to the damage of the contact.

Material	**Electrical Conductivity %IACS**	**Melt Voltage**	**Arc Voltage**	**Arc Current**
Cadmium	24	-	10	0.5
Copper	100	0.43	13	0.43
Gold	77	0.43	15	0.38
Nickle	25	0.65	14	0.5
Palladium	16	0.57	15	0.5
Silver, fine	105	0.37	12	0.4
Tungsten	31	0.75	15	1.0

Table 2.3. Characteristics of Various Contact Materials (Slade, 2013)

Contact life is deemed to have reached failure when the contacts stick or weld together, or if excessive material transfer has taken place to either one of both contacts and a good electrical contact make is no longer possible. These failure modes are due to successive switching operations and of material loss due to splattering.

The material transfer takes place as a result of joule heating. As the contact area separates, the area of the contacts diminishes. The load current is then forced to flow through an ever more constricted area, and this causes a build-up of heat, which reaches such a point where the contact material is melted and then boils. With a dc

load, this liquefied material tends to deposit on the cathode of the contact, simply due to the fact that it is cooler than the anode. Material transfer also occurs as a result of arcing, with the transfer being opposite to above and depositing the molten metal on the anode of the contact.

Material loss due to boiling and arcing is from vaporization and splattering respectively, during contact bounce on the closure of the contacts. Although the amount of material loss is minuscule, over tens or hundreds of thousands of operations it becomes significant.

2.5.1 Contact Bouncing

The making of the contacts is not usually finished at first touch, but as a consequence of bouncing the members make and break their contact several times before they reach a permanent state of contact. This can have implications due to the many disturbances bouncing brings. The exactitude of contact make is lost, and the material transfer by arcs and bridges is increased, since each bounce is the same as a new switch operation. A contact is particularly vulnerable to damage by re-bounce when the current begins with a high inrush as in the case of inductive loads, such loads may result in current in excess of eight times the normal operating current. (Mcbride, 1989) & (Mcbride, 1991).

2.5.2 The Formation of the Electric Arc

The formation of electric arcs can happen both when the contacts are opening and closing. The more destructive of the two scenarios, is when the contacts are opening and this will be addressed first.

2.5.3 The Formation of the Electric Arc during the Opening of Contacts

An arc will always form between opening contacts if the circuit current and the voltage that appears across the contacts is greater than a minimum value. The arc formation depends entirely upon the properties of the contact material and the arc always initiates in the metal vapour from the contacts themselves (Slade, 2013).

From equation (5) the contact resistance R_C is given by

$$R_c = \frac{\rho}{2a} = \frac{\rho}{2}\sqrt{\frac{\pi H}{F}} \qquad (21)$$

where again, a is the radius of the real area of contact, H is the material or Vickers hardness, ρ is the resistivity and F is the force holding the contacts together.

As the contacts separate, the holding force F → 0 and therefore the area of contact a → 0, producing an increase in the contact resistance R_c. This increase in the contact resistance R_c leads to the voltage drop $U_c$across the contact also increasing.

$$U_c = IR \tag{22}$$

From 2.4.3 above the relationship between voltage and temperature has been discussed, an approximation to this can be represented by the temperature of the contact T_c being given by

$$T_c^2 = T_0^2 + U_c^2 \times 10^7 \text{ K} \tag{23}$$

where T_0 is the ambient temperature. Work carried out by (Wakatsuki, 2008) shows a stage will be reached when the temperature of the contact spot will equal to the melting point T_m of the contact material.

The values from table 2.4 below and calculated values from equation (23) are depicted in figure 2.6 below for a wide range of contact materials and shows the correlation between results.

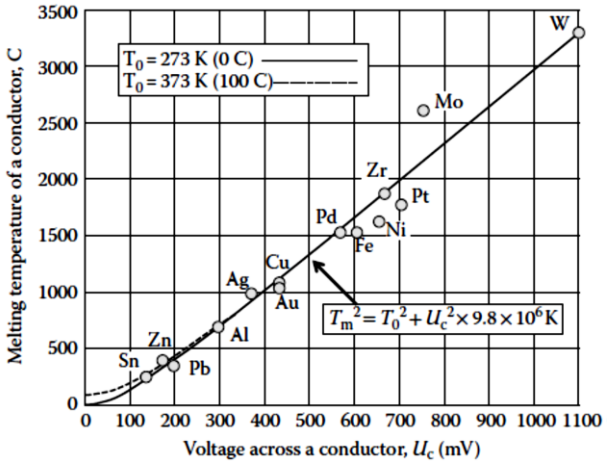


Figure 2.6. Showing the relationship between the calculated and measured melting temperature and the measure voltage drop across the contact (Wakatsuki, 2008).

When a contact reaches this melting stage and the contact pairs continue to draw apart, a molten bridge is formed between them, this happens at low currents (Utsumi, 1969), (Miyajima, 1998) and (Ishida, 2004), slow (Mcbride, 2012) and high speed opening contacts (Slade, 1971), (Koren, 1975) and (Slade, 1972) and even in a vacuum (Slade, 2008). This molten bridge continues to form and is drawn until it ruptures due to instability. After this rupture of the molten metal bridge, an arc forms

in its vicinity. A typical change in the initial voltage drop across the contacts is shown in Figure 2.7 (Slade, 2013).

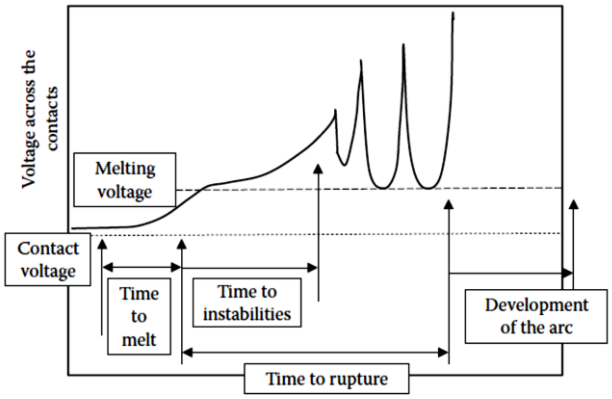


Figure 2.7. Showing the change in the initial voltage drop across the contacts as the arc develops (Slade. 2013).

These voltage characteristics can be described using the four stages shown in figure 2.8 (Haug, 1990) & (Slade, 2010):

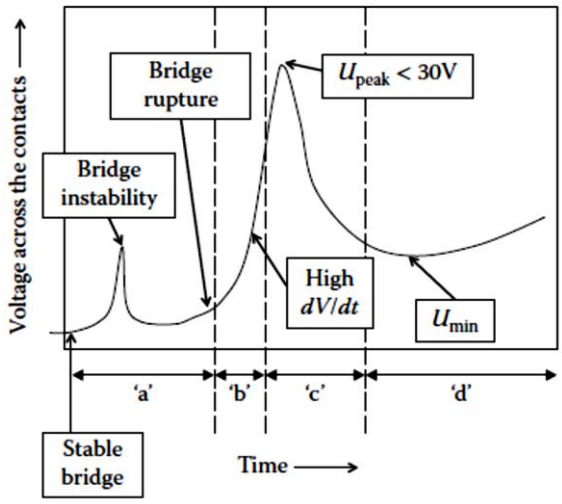


Figure 2.8. The four stages of molten metal bridge rupture and metal phase arc formation.

Stage (a): Once the molten metal bridge has formed its rate of change of voltage is about 2×10^3 Vs^{-1}. As the contacts continue to open and the bridge is drawn further it becomes unstable. There are a number of physical reasons for this instability, including surface tension effects, boiling of the highest temperature region, convective flows of molten metal resulting from the temperature variation between the bridge roots and the high-temperature region.

The bridge will eventually rupture, releasing metal vapour into the contact gap when the voltage across it, U_b, is close to the calculated boiling voltage U_{b1} of the contact materials, i.e.:

$$U_b = \sqrt{L(T_{b1}^2 + T_0^2)} \qquad (24)$$

where L is the Lorenz constant and T_{b1} K is the boiling temperature as in the table 2.4 below.

Metal	Breaking Voltage(U_b)	Calculated Boiling Voltage(U_{b1})	Boiling Temperature(T_{b1}) K
Ag	0.75	0.77	2485
Cu	0.8	0.89	2870
W	1.7	1.82	5800
Au	0.9	0.97	3090
Ni	1.2	0.97	3140
Sn	0.7	0.87	2780

Table 2.4. Showing the Boiling Voltage (U_{b1}) with the Break Voltage (U_b) for Various Metals (Slade, 2013)

Stage (b): Once the bridge ruptures the voltage across the contacts rises very rapidly without a discontinuity from about 10^3 Vs^{-1} to about 10^9 Vs^{-1}

This rate of rise of the voltage will depend upon the dimensions of the molten metal bridge just before its rupture. After the bridge rupture a very high pressure, perhaps as high as 100 atmospheres (Haug, 1990) & (Slade, 2010), very low electrical conductivity, metal vapour exists between the contacts. This region can then be considered to be a capacitor with a very small capacitance. Because the circuit's inductance prevents a rapid change in current charge flows from the circuit inductance into this small capacitor causing the very high dV/dt. The metal vapour volume expands rapidly into the surrounding lower pressure ambient and as it does its pressure also decreases rapidly. When the pressure of the metal vapour decreases to 3–6 atmospheres conduction is initiated with a voltage across the contacts of a few 10's of volts.

At these pressures the discharge that forms is the "pseudo arc" (Puchkarev, 1997), (Ebling, 1991) where the current is conducted by ions. During this stage the electrons required for charge neutrality will be introduced into the discharge from secondary emission resulting from ion impact at the cathode. As the original molten metal bridge will have material from both the cathode and the anode, net transfer of material from the anode to the cathode is expected and is indeed observed (Haug, 1990) & (Slade, 2010).

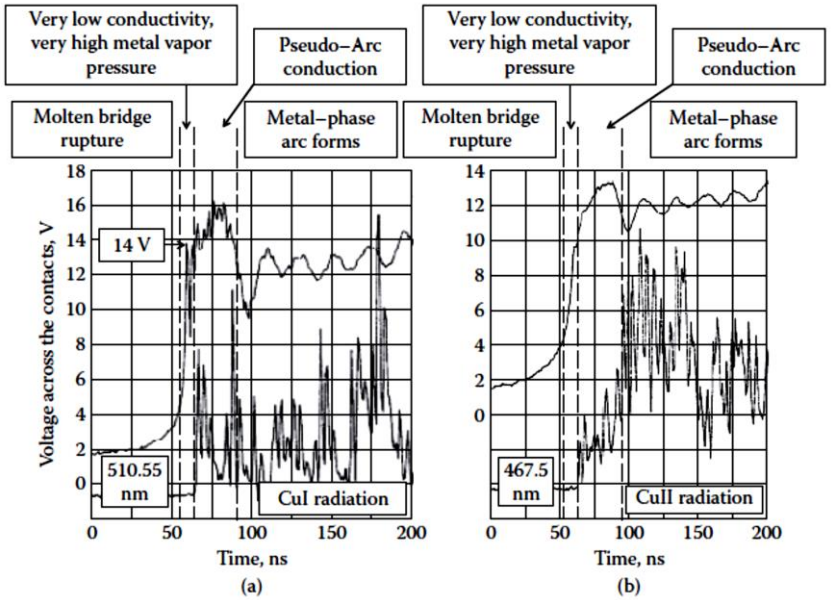


Figure 2.9. The transition from molten metal bridge to the metal phase arc (Haug, 1990): (a) Voltage, neutral copper radiation (Cu I), Cu contacts, 50 A, 100 V and (b) voltage, ionized copper radiation (Cu II), Cu contacts, 50 A, 100 V.

Stage (c): As the pressure of the metal vapour continues to decrease to about 1–2 atmospheres, the pseudo arc transitions into the usual arc discharge with an arc voltage impressed across the contacts whose value is about that of the minimum arc voltage expected for an arc operating in the contacts' metal vapour (i.e., $U_{min} \approx 10 - 20$ V). Here again net material transfer will be from anode to cathode. It is only in a vacuum ambient that the arc continues to operate in metal vapour evaporated from the contacts themselves (Slade 1972), (Slade, 2008) & (Slade, 2008b). In order to sustain this arc a minimum arc current is also required.

Stage (d): At this stage as the contacts continue to open and the arc between them gradually transitions from the metallic phase arc to the ambient, gaseous phase arc with most of the current now carried by electrons. The whole sequence is illustrated in Figure 2.10 (Slade, 2013). As the contacts continue to open this metallic phase arc transitions into an arc operating in the ambient atmosphere.

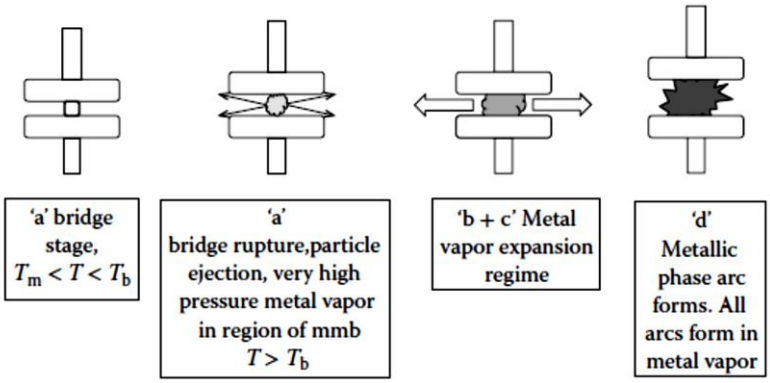


Figure 2.10. Showing the opening sequence of an electrical contact; the formation of the molten bridge; its rupture and arc formation (Slade 2010).

2.6 Arcing in a DC Circuit and Material Transfer

One of the most important consequences of arcing is the effect that the arc has on the erosion of the contact material. The erosion of contacts occurs because both the cathode and the anode under the roots of the stationary arcs can be heated to the boiling point of the contact material. Even when the arc moves rapidly across a contact surface the arc roots still melt the contact surface directly under them. (Slade, 2013) summarises the erosion parameters. As the amount of erosion per operation of the contact depends upon many parameters as shown below:

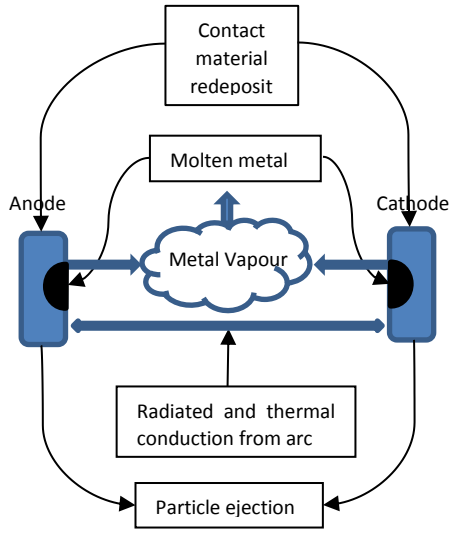


Fig 2.11. Energy balance at the contacts from the arc and the effect on contact erosion.

1. The circuit current
2. The circuit resistance, inductance and capacitance
3. The arcing time
4. The contact material
5. The structure of the contact material

6. The shape and size of the contact
7. The contact's attachment
8. The opening velocity of the contact
9. The bounce on making contact
10. The open contact gap
11. Motion of the arc on the contacts
12. The design of the arc chamber
(a) Gas flow
(b) Insulating materials used

Contact erosion is further complicated by mechanical stresses on the contact as a result of the impact on closing. These stresses caused by the switch mechanism affect the contact materials in widely different ways which depend upon the material and the manufacturing process. In principle the mass lost per operation of the contact should be given by

Mass loss = *f* (total power input into the contacts) (Slade, 2013).

However, both sides of this equation present complexities that prevent it from being established let alone solved. For example, Slade asks the question what is meant by the mass loss. The total mass loss from a contact is a mixture of the following components:

• Metal vapor evaporated from the arc roots +
• Metal droplets ejected from the arc roots –
• Metal re-deposited back onto the contact faces –
• Metal deposited from the opposite contact

The actual erosion products are difficult to predict from a given switching event even when the parameters are carefully controlled (Slade, 2008), (Puchkarev,1997) & (Ebling, 1991)

The modeling of the total power input into the contacts also presents problems. First of all, for most designers of switches, it is only possible to measure the current in the circuit $I(t)$ and also perhaps the arc voltage $U_A(t)$across the contacts during the arcing. If we take some of the power input components at the cathode,

• Power input from ions = $f_1(I \times Cathode\ fall\ voltage)$
• Power input from radiation = $f_2(I \times U_A)$

• Power input from neutral atoms = $f_3(I \times U_A)$

• Joule-heating of contact = $f_4(I^2)$

As well as the mass loss, calculation of the total power input into the contacts can be difficult, in terms of measurement of arc voltage U_A and circuit current I_C.

Hence, calculation of contact erosion is still a topic of research, there is a great deal of literature on contact erosion, but it tends to be application specific and subject to guidance when used for design.

Conclusion

This chapter started by exploring the constituent components that make up the relay and put together a framework which enables failure modes to be clearly identified.

This chapter provides the following conclusions

- Electrical wear of the contacts was identified as the primary mode of failure, due to Joule heating and erosion of the surface to arcing, contamination and corrosion.
- The physics of electrical contacts was examined from reviewing present literature and the constriction (contact) resistance was deemed to be an important metric in determining how the condition of the contact may be determined, as well as its importance in the development of failure. Joule heating was explored, as the relationship between temperature and constriction resistance and voltage was identified via the Wiedeman-Franz law.
- Lastly the formation of an arc between the contacts was discussed and how this erodes the surface of the contact due to material loss and mass transfer, this is scenario dependant. It was noted a degree of arcing can be useful to remove oxides and film that collect on the contacts of the relay, but excessive arcing causes reduced life and whereas arcing suppression is recommended by manufactures (CR and snubber circuitry), it cannot be eliminated altogether and influences the prediction of how long the relay contacts will last.

Chapter 3 – Condition Based Monitoring (CBM) and Prognostics Literature Review

As the Thesis is concerned with the implementation of Condition Based Monitoring (CBM) and Prognostics, the primary aim of this chapter is to provide a detailed literature review of these topics, along with a review of maintenance strategies. The various prognostics approaches are examined in detail and initial conclusions are drawn on the suitability of methodologies for use in the application of relays.

3.1 Maintenance Evolution

Traditionally, maintenance has fallen into two philosophies, namely

1. Reactive maintenance (sometimes called unplanned)
    - Whereby corrective maintenance is carried out when a fault occurs
2. Proactive maintenance (sometimes called pre-planned)
    - Whereby preventative maintenance in undertaken
    - Also predictive maintenance is employed

Until the advent of the Second World War, machinery was relatively simple in terms of control and instrumentation, slow running, with a rugged build quality. Production demands were not so severe, thus downtime due to breakdown was not so detrimental, and hence machinery was maintained on a breakdown by breakdown basis.

After the war, with the rebuilding of Germany and Japan, a much more competitive marketplace was formed and an intolerance of downtime became much more apparent. Labour costs grew, resulting in a greater attraction to mechanisation and automation. Machinery too, was lighter in construction, as well as being greater in complexity with higher running speeds, resulting in a perceived reduction in reliability. The solution to this was to implement proactive maintenance (Brown and Sondalini, 2014).

Preventative maintenance is one such sub-group of proactive maintenance, where maintenance is performed on a periodical basis. Service periods are determined from the use of past historical data based on Mean Time Between Failures (MTBF) and not from any input from the machinery itself. The machinery is thus serviced at the

scheduled period, whether it is necessary or not. This sort of maintenance is still adopted, with routine examinations, replacement of lubrications, filters and component overhauls. Typical examples are in automotive and aviation. However, reactive and preventative maintenance both have safety and financial implications associated with them.

The 1960s resulted in the aviation industry carrying out effectiveness studies on the then current maintenance strategies which assumed “the older equipment gets the more likely it is to fail”. The results apportioned out of the six failure modes identified, only three modes where attributed to age related failure and of these three modes, 11% were age related. This highlighted an open ended, constant probability of failure, but however by monitoring the change of a suitable parameter, gave the ability to monitor a change in condition. The study resulted in massive reduction in maintenance hours (30:1), but equally there was a dramatic increase in safety as well. This approach led to the foundations of Reliability Centred Maintenance (RCM), RCM performs two tasks: firstly, to analyse and categorise failure modes (e.g. FMEA) and secondly, assess the impact of maintenance schedules on system reliability (Kothamasu et al., 2006).

The 1980s’ heralded more increasingly complex plant and systems being introduced, the competiveness of competing marketplaces meant downtime became more intolerable and added to this, maintenance costs based on preventative maintenance began to spiral. Industry looked to a more predictive maintenance methodology, RCM employing manual inspections and simple data trending, awareness of failure processes, improved management techniques and new technologies allowed greater understanding of machine and component health.

The 1990s’ saw maintenance strategies starting to embrace other issues such as risk, with environmental and safety becoming paramount. New concepts had been developed; expert systems, just in time manufacturing, quality standards and condition based monitoring (CBM) which will be discussed in more detail later in this chapter.

New concepts were proposed in the 21st century; Prognostics, Integrated Vehicle Health Management (IVHM) and Integrated System Health Monitoring (ISHM) are all developmental techniques that are well founded in literature.

In conclusion, system reliability and availability, with minimised downtime and failure are the goals for many industries. Maintenance strategies have moved from reactive to a proactive stance, as systems and equipment become more complex and expensive, and increasing competition drives industries to become more lean and efficient, industrial and military communities are becoming increasingly concerned about system reliability and availability (Vachtsevanos et al., 2006). In many industries using complex machinery, the need to reduce maintenance costs, minimise the risk of catastrophic failures, and maximise system availability is leading a drive toward a new maintenance philosophy.

3.2 Integrated Vehicle Health Management (IVHM)

Integrated Vehicle Health Management (IVHM) or Integrated System Health Management (ISHM) is the unified capability of systems to assess the current or future state of the member systems health and integrate that picture of system health within a framework of available resources and operational demand (Jennions, 2011).

IVHM was born as a concept from taking the proven aviation maintenance methods already in existence and forms the next step from condition based maintenance. As technology such as sensors and data processing improved as well as the understanding of the systems concerned grew, it became possible to not just detect failure but also to predict it. IVHM is particularly attractive in high unit cost and high maintenance cost platforms such as aircraft and spacecraft. NASA was one of the first organisations to use the name IVHM to describe how they wanted to approach maintenance of spacecraft in the future.

(Jennions, 2011) documents the generic IVHM taxonomy consisting of following sub-categories:

- Maintenance service offerings (e.g. CBM, Total Care, RCM)
- Business (e.g. Business models, IVHM mapping)
- System design
- Architecture
- Analytics (e.g. Diagnostics, Prognostics)
- Technologies (e.g. Structural Health Management (SHM))

IVHM enables many disciplines with an integrated framework. CBM, Health and Usage Monitoring Systems (HUMS), and RCM are some of the maintenance strategies offered under IVHM, where diagnostics and prognostics are considered under the analytics category.

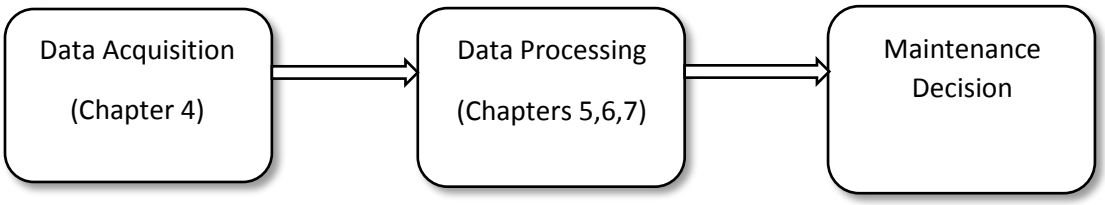
This Thesis is concerned with the analytics shown in the framework above, in particular, the application of prognostics to systems and develops solutions to fulfil this.

3.3 Condition Based Monitoring (CBM)

CBM is a maintenance program that recommends maintenance actions based on the information collected through condition monitoring and forms a key component of the IVHM framework. CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviours of a physical asset. A CBM program, if properly established and effectively implemented, can significantly reduce maintenance cost by reducing the number of unnecessary scheduled preventive maintenance operations, (Jardine et al., 2006).

A CBM program consists of three key steps, (Lee et al., 2004).

1. Data Acquisition step (information collecting), to obtain data relevant to system health.
2. Data Processing step (information handling), to handle and analyse the data or signals collected in step 1 for better understanding and interpretation of the data.
3. Maintenance Decision-Making step (decision-making), to recommend efficient maintenance policies.



3.3.1 Data Acquisition

The data acquisition step involves collecting and storing useful information about the asset that will enable CBM. There are two main classes of data that are collected, event based and condition based data. The event based data may contain information

that would be found in logs, e.g., when the machine/component was installed, breakdowns, overhauls, causes, and/or what was done to rectify the problem (e.g. preventative maintenance, lubricant changes, replacement parts, modifications etc.). Condition monitoring data is measurement based and relates to ascertaining the health condition/state of the physical asset.

Event data is usually entered into the system manually once a maintenance action has been completed, and Computerised Maintenance Management Systems (CMMS), (Davies, 2000), have been developed to handle the management of such data.

The types of measurements that may be used to gather condition based monitoring data is based on largely sensor availability, the range of data that may be monitored is ever growing as sensor technology develops. As well as parameters such as vibration, acoustics, oil analysis, environmental, ultrasonic etc., new sensors are being developed to collect data such as vibration inside a gas turbine whilst in use (Kirianaki, 2002), (Senesky, 2009).

Wireless technologies, such as Bluetooth, have provided an alternative solution to cost-effective data communication. There may also be constraints to how condition based data is collected, the Civil Aviation Authority (CAA) for example, dictate that measurements must be non-invasive.

3.3.2 Data processing

The data that is collected will need to be cleaned. Data cleaning is important for both event based and conditioned based data. Data entered manually, which is usually the case with event based data, is often subject to errors as it involves a human element. Condition based data is subject to measurements from instrumentation and will incur a degree of measurement error; sensor faults and interference can also cause data error.

The next step of data processing is data analysis. A variety of models, algorithms and tools are available in the literature to analyse data for better understanding and interpretation of data. The models, algorithms and tools used for data analysis depend mainly on the types of data collected. (Jardine et al., 2006) state condition monitoring data collected from the data acquisition step are versatile. It falls into three categories:

Value type: Data collected at a specific time epoch for a condition monitoring variable are a single value. For example, oil analysis data, temperature, pressure and humidity are all value type data.

Waveform type: Data collected at a specific time epoch for a condition monitoring variable are a time series, which is often called time waveform. For example, vibration data and acoustic data are waveform type.

Multidimensional type: Data collected at a specific time epoch for a condition monitoring variable are multidimensional. The most common multidimensional data are image data such as infrared thermographs, X-ray images, visual images, etc.

The way each type of data set is processed is the subject of many papers in literature, a variety of tools, algorithms and models are available to deal with the interpretation and extraction of features from each type of raw data. Waveforms and multidimensional data are subject to signal processing techniques.

Multidimensional data such as raw images are usually very complicated and immediate information for fault detection is unavailable. In these cases, image processing techniques are powerful tools to extract useful features from raw images for fault diagnosis (Nixon, 2002). Examples of where image processing techniques have been used for condition based monitoring and fault diagnosis are (Wang et al., 1993) and (Utsumi et al., 2001).

For waveform data analysis there are three main categories of analysis; time-domain, frequency-domain and time-frequency domain analysis.

Time-frequency domain methods are based upon the time waveform itself and include analysing the waveform in term of its descriptive statistics. Mean, standard deviation, peak-to-peak interval, crest factor as well as higher order statistical characterisations such as root mean square, kurtosis and skewness.

Time series modelling is an approach that is based on fitting the waveform data to a parametric time series model. Popular models in literature are Auto Regressive (AR) and Auto Regressive Moving Average (ARMA). (Isermann, 2011), (Pham, 2010) & (Carden, 2004).

Frequency-domain analysis is based on the transformed signal in frequency domain.

The advantage of frequency-domain analysis over time-domain analysis is its ability to easily identify and isolate certain frequency components of interest. Fourier transform spectrum analysis techniques by means of a Fast Fourier Transform (FFT) are used to look at either components of interests within a frequency band or the whole spectrum and extract features of interests. (Lu et al., 2009), (Hameed et al., 2009), (Nandi et al., 2005).

The Cepstrum, defined as the power spectrum of the logarithm of the power spectrum has the capability to detect harmonics and sideband patterns in power spectrum. High-order spectrum, i.e., bi-spectrum or tri-spectrum, can provide more diagnostic information than power spectrum for non-Gaussian signals.

There are two main classes of approaches for power spectrum estimation, namely the non-parametric approaches that estimate the autocorrelation sequence of the signal and then applies a Fourier transform to the estimated autocorrelation sequence. Secondly, the parametric approach that build a parametric model for the signal and then estimate power spectrum based on the fitted model. Among them, AR spectrum (Dron, 1998), (Stack, 2004) and ARMA spectrum (Salami, 2001) based on AR model and ARMA model, respectively, are the two most commonly used parametric spectra in machinery fault diagnostics.

Time-frequency analysis may a have significant advantage if the signal being monitored is non-stationary, which is common in moving machinery faults. Time-frequency analysis has the ability to analyse waveforms in both the time and the frequency domain and has been developed for non-stationary signal investigation. Traditional time–frequency analysis uses time–frequency distributions, which represent the energy or power of waveform signals in two-dimensional functions of both time and frequency to better reveal fault patterns for more accurate diagnostics. Examples include spectrograms and Wigner–Ville distributions.

Wavelets present a time-frequency analysis transform, which differs to the time-frequency approach by using a time-scale representation of a signal. Wavelets have been applied across numerous condition monitoring applications (Peng, 2004), (Zhu, 2009) and (Watson et al., 2010).

Traditionally, reliability analysis was carried out by fitting the event data to a time between events by a probability distribution, then this distribution was used for further analysis. CBM gives additional data and it is beneficial to use both the event and condition monitoring data together. From this combined analysis mathematical models can be built that properly describes the underlying mechanism of a fault or failure. The models built based on both event and condition monitoring data is the basis for maintenance decision support—diagnostics and prognostics (Jardine et al., 2006).

Models include time-dependent Proportional Hazards Models (PHM); a commonly used parametric baseline hazard function is the Weibull hazard function, which is the hazard function of the Weibull distribution.

The concept know as potential-to-functional failure interval (P-F) is used in (RCM), the P-F describes failure patterns in condition monitoring. The P-F, is the time between a potential failure (P), which is some indicator of condition and the actual functional failure (F). Although difficult to quantify in real applications, the P-F interval offers a useful metric in condition monitoring.

The Hidden Markov model (HMM), (Rabiner, 1989), (Elliot, 1995), is another appropriate model for analysing event and condition monitoring data together. An HMM consists of two stochastic processes: a Markov chain with finite number of states describing an underlying mechanism and an observation process depending on the hidden state. A discrete-time HMM is defined by

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}\boldsymbol{X}_k + V_{k+1} \quad (25)$$

$$\boldsymbol{Y}_k = \boldsymbol{C}\boldsymbol{X}_k + W_k \quad (26)$$

where $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ denote the hidden process and the observation process, respectively, V_k and W_k are noise terms with martingale increments, and A and C are parameters. Event data and condition monitoring data are used to train the HMM, i.e., to estimate model parameters. (Bunks et al., 2000) and (Lee et al., 2004)

3.4 Maintenance Decision Support

Within the maintenance decision making step, there are two important aspects in a CBM program, namely, diagnostics and prognostics. Diagnostics deals with fault detection, isolation and identification, when it occurs. Fault detection is a task to indicate whether something is going wrong in the monitored system; fault isolation is a task to locate the component that is faulty; and fault identification is a task to determine the nature of the fault when it is detected. Up to now, diagnosis has been used within safety critical equipment to alert the operator of a fault with the electromagnetic relay e.g. operational failure, via systems such as Built in Test (BIT) as used in the FADEC.

Prognostics deals with fault prediction before it occurs. Fault prediction is a task to determine whether a fault is impending and estimate how soon and how likely a fault will occur. Diagnostics is posterior event analysis and prognostics is prior event analysis. Prognostics is much more efficient than diagnostics to achieve zero-downtime performance. After a brief look at diagnostics, prognostics will be the subject of the rest of this chapter.

3.4.1 Diagnostics

Diagnostics, is required when fault prediction of prognostics fails and a fault occurs. A CBM program can be used to carry out diagnostics or prognostics, or both. No matter what the objective of a CBM program is, however, the three CBM steps are followed as in the diagram.

Fault diagnostics is primarily involved with the mapping of the information gained from the machine measurements or features in the features space to the actual machine faults in the fault space. This process is also called pattern recognition (Jardine et al., 2006). In the past this was a skilled task carried out by experience personnel utilising the data gained from data processing. As the development of machine learning techniques has become prevalent and reliable over the last few years, this is increasingly being done automatically using statistical approaches and artificial intelligent approaches.

3.4.2 Statistical Methods

Statistical methods attempt to determine whether a specific fault is present or not based upon the data from condition monitoring, and typically uses hypothesis testing to indicate if a fault is present. Statistical distributions or specific models of the fault are tested against distributions or models of the test data and the hypothesis is either true or false (Sohn et al., 2002).

In another approach, Statistical Process Control (SPC) has been widely used in condition monitoring for fault detection and diagnostics. The deviation from a reference signal which represents the normal operation is monitored. Boundaries are then used to determine if the measured signal deviates from the reference signal by a certain amount that determines impending failure. The technique has been used by

(Lee, 2004) for individual component identification and (Fugate, 2001) for vibration based damage detection.

Cluster analysis, as a multivariate statistical analysis method, is a statistical classification approach that groups signals into different fault categories on the basis of the similarity of the characteristics or features they possess. It seeks to minimise within-group variance and maximise between-group variance. The result of cluster analysis is a number of heterogeneous groups with homogeneous contents: There are substantial differences between the groups, but the signals within a single group are similar (Jardine et al., 2006). Distance measurements are commonly used to group signals based on the similarity between two signals (Skormin, 1999), (Artes, 2003). Euclidean distance, Mahalanobis distance, Kullback–Leibler distance and Bayesian distance are all measurements that have been explored in literature.

3.4.3 Artificial Intelligence

Artificial Intelligence techniques are increasingly being used, and the results in literature have shown good performance. AI require training and specific knowledge which can be hard to obtain. Artificial Neural Networks (ANNs), Expert Systems (ES), Fuzzy Logic, Fuzzy Neural Networks, Neural-Fuzzy systems and evolutionary algorithms (EAs) have all been used, (Rafiee, et al 2007) & (Saxena, 2007).

3.4.4 Model-Based Approaches

Model-based approaches for machine fault diagnosis have also been utilised. These approaches use the physical attributes to produce an explicit mathematical model of the equipment being monitored. This model may then be used with residual generation methods such as parameter estimation (or system identification), Kalman Filtering and parity relations to generate signals, called residuals, which are indicative of fault presence in the machine (Jardine et al., 2006). Finally, the residuals are evaluated to arrive at fault detection, isolation and identification. (Simani, 2003).

3.5 Prognostics

Unlike diagnostics, which has started to become an established practise, prognostics is still a large area of research and is subject to changing standards. Within literature, different definitions of prognostics include; “prognostics is, or should be, performed at

the component or sub-component level"; "prognostics involves predicting the time progression of a specific failure mode from its incipience to the time of component failure"; "an appreciation of future component operation is required"; "prognostics is related to, but not the same as, diagnostics." (Sikorska et al., 2011).

A definition by the ISO 13381-1 standard gives the definition of prognostics as 'an estimation of time to failure and risk for one or more existing and future failure modes', (ISO 13381-1, 2004). Certainly as the technologies mature, the definitions will become more standardised, however, at present the majority of work done within the field is of a theoretical nature, with few published examples of prognostic models being applied in real scenarios, under a normal range of operating conditions and how this may impact on business models.

Without doubt, the potential advantages of prognostics is great, compared with just using diagnostics as part of a CBM system; as only the downtime due to the actual maintenance action becomes relevant. The ability to plan and schedule maintenance whilst the system is up and running offers significant savings in cost, logistics, maintenance downtime and life cycle costs. System design and development, reliability, safety can also be added to this list (Sun, Bo, et al., 2010). Example of these potential cost savings are given by (Hecht, 2006) for aviation assets, where maintenance and re-test can be pre-planned prior, yielding a saving of maintainer time and significant reduction in its variability.

In published literature that relates to prognostics, it can be seen that there is a strong correlation between and the high reliance on diagnostics. However the boundary between the two is somewhat blurred. Certainly the damage that has occurred needs to be identified and quantified as a starting point for applying prognostics, this is a retrospective activity, while prognostics is concerned with trying to predict the damage that is yet to occur. Although diagnostics may provide useful business outputs on its own, prognostics relies on diagnostic outputs (e.g. fault indicators, degradation rates etc.) and therefore cannot be done in isolation. (Sikorska et al., 2011). Detection of failure progression is more valuable compared to the detection of failure once it has reached a severe point. Furthermore, it is a prerequisite for prognostics (Xiong et al., 2008).

The process of prognostics involves two main stages, firstly the current health status of the system/component must be determined and this stage may considered under the heading of diagnostics. Other terms used in literature to describe this stage are; degradation identification and health assessment. Various methods may be used to fulfil this task, but as discussed above in section 3.4 the process tends to be carried out automatously, mainly by Bayesian methods and AI, utilising such techniques as pattern recognition by clustering and classification. Both of these techniques have advantages and disadvantage associated with them. The former generally requiring a model of the degradation, this involves an understanding of the physical process that accounts for the degradation and can be very difficult, if not almost impossible to determine. Where in the latter techniques, to infer an accurate degradation pattern model from an A.I based approach can utilise a large amount of failure data. This data may be difficult to obtain in large quantities, as in general, systems are not run to failure and in some systems this may take months/years to evolve.

The second stage of the prognostic is the estimation of the RUL by prediction of the degradation trend. Prognostics implies forecasting of the systems/components future health level by propagating the current health level until a failure threshold. Terms used for describing this phase are extrapolation, propagation of fault, trending, tracking and time series analysis.

The range of possible prognostic approaches and their applicability for different system in terms of increasing accuracy and cost is shown in figure 3.1 below (Vachtsevanos et al., 2006). It can be seen that experienced based prognostics is the cheapest option and its range of system applicability is diverse, consisting of statistical (e.g. Weibull) and knowledge based models (e.g. fault log data), however in terms of accuracy, it falls short of Physics of Failure (PoF) and Data Driven techniques which form the middle section of the pyramid. Model-based prognostics, at the top of the pyramid, are the most accurate but difficult and costly to develop, but are expected to take prognostics into the future.

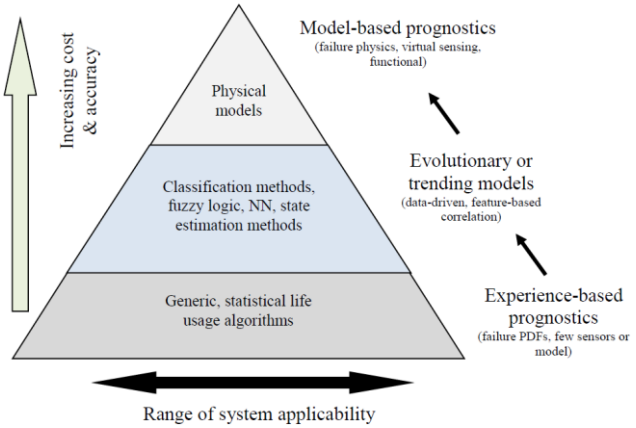


Figure 3.1. Possible prognostic approaches and their applicability for different system in terms of increasing accuracy and cost (Vachtsevanos et al., 2006)

Prognostic technologies typically use measured or inferred features, in combination with data-driven and/or physics-based models, to predict the condition of the system at some future time. Prognostic techniques combining data-driven and physics-based models are sometimes referred to as 'hybrid' prognostics.

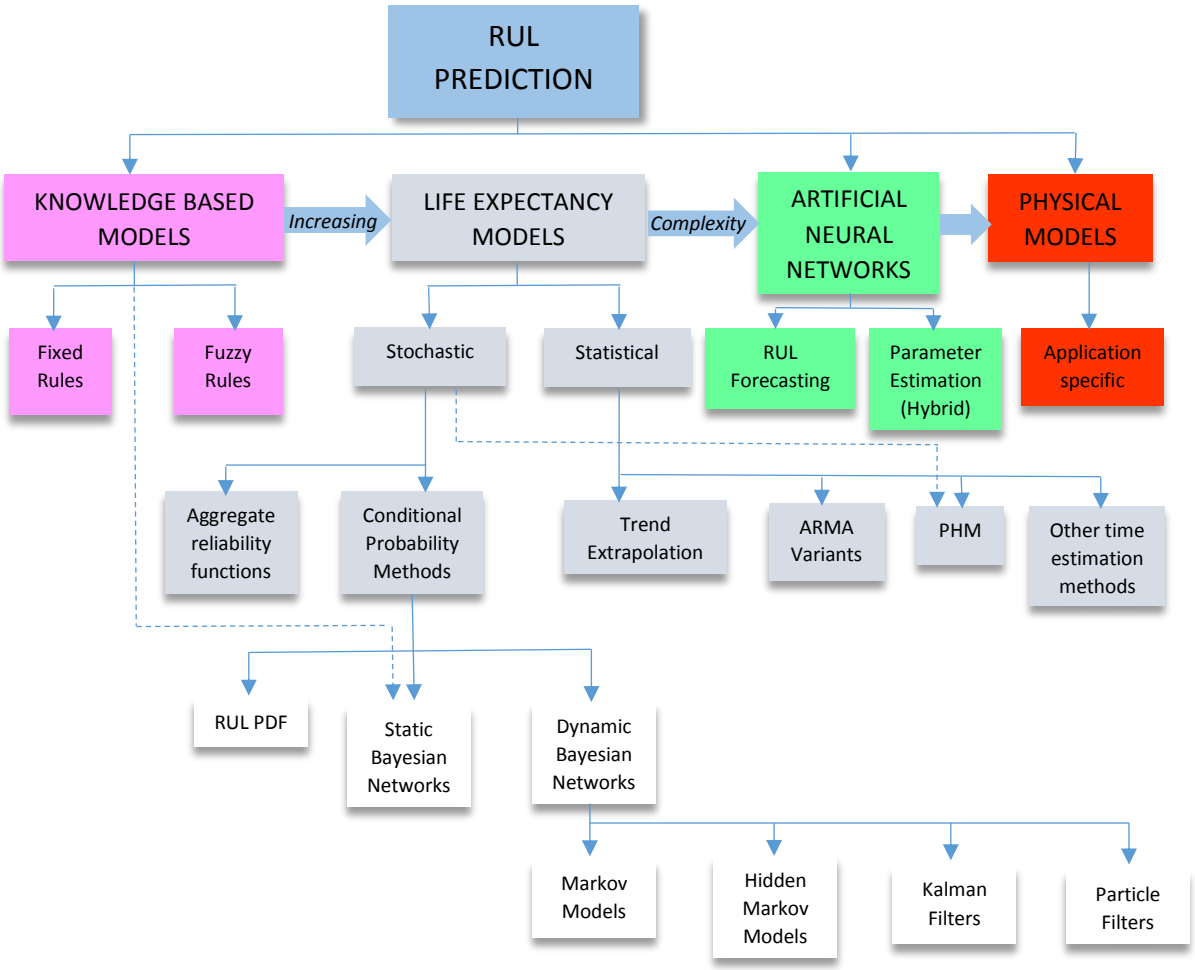


Figure 3.2. Main categorisation of model for RUL prediction (Sikorska et al., 2011),

In the figure below, the main categories of prognostics techniques for remaining useful life prediction used in literature are presented. The next section will look at the majority of these techniques in the form of literature review to enable the best solution for the problem to be proposed.

3.6 Review of Prognostic Techniques

3.6.1 Model Based Prognostic Techniques

Model-based prognostic schemes include those that employ a dynamic model of the process being predicted. These can include physics-based models, autoregressive moving-average (ARMA) techniques, Kalman/Particle filtering, and empirical-based methods.

In engineering, the development of a model has been used traditionally to understand component failure mode progression. The development of a Physics of Failure (PoF) or Physics based models (PbM), (the two names will be used interchangeably throughout this Thesis), that incorporates the ability to assess damage to a component, taking into account operating conditions and gives a cumulative damage assessment provides a basis to evaluate the distribution of RUL. The results from such model may be used as a basis for real-time failure prognostic predictions with specified confidence bounds. These confidence bounds may be the result of statistical representations of historical operational profiles.

The actual availability of physics of failure models is limited. A model that incorporates the ability to predict damage parameters with an acceptable confidence boundary over steady-state and transient loads, temperatures and other variables are still quite few. PoF models are implemented in three different ways (Sikorska et al., 2011); firstly, dynamic ordinary or partial differential equations that can be solved with approximation approaches (e.g. Lagrangian or Hamiltonian dynamics), secondly, state-space methods (systems of first order differential equations) and thirdly, simulation methods. One such model that has found prevalence in prognostic literature is the Paris Law for crack propagation (Zhao et al, 2013), (Zhao et al, 2015).

$$\frac{d\alpha}{dN} = C_0(\Delta K)^n \tag{27}$$

where α = instantaneous length of dominant crack

N = running cycles

C_0, n = material dependant constants

ΔK = range of stress-intensity factor over one loading cycle

The crack growth process was simulated to yield normally distributed crack lengths, which were used within a Bayesian framework to update the parameters of the degradation model (e.g., Paris' law). The degradation model was initially fed by the results of a stress analysis from a gear dynamic model or finite element model. The distributions of the uncertainty factors were updated via Bayesian Inference using the condition monitoring data (simulated crack lengths), and an estimation of the RUL based on the degradation model was provided.

In terms of electronics, (Kulkarni, Chetan S., et al., 2012) proposes first principles based modeling and prognostics approach for electrolytic capacitors. Electrolytic capacitors and MOSFETs are the two major components, which cause degradations and failures in DC-DC converters. The paper studies the effects of accelerated ageing due to thermal stress on sets of capacitors, with the focus on deriving first principles degradation models for thermal stress conditions. The degradation data forms the basis for developing the model based remaining life prediction algorithm. Finally the data is used to derive accurate models of capacitor degradation, and use them to predict performance changes in DC-DC converters.

(Fan,et al., 2011) looked at how to accurately predict the reliability of LED lighting. In this paper, after analyzing the materials and geometries for high-power white LED lighting at all levels, i.e., chip, package and system; Failure Mode Mechanisms and Effects Analysis (FMMEA) was used in the PoF-based PHM approach to identify and rank the potential failures emerging from the design process. The second step in this paper was to establish the appropriate PoF-based damage models for identified failure mechanisms that carry a high risk.

A PoF based prognostic method for power electronic modules was proposed by (Yin, C. Y., et al., 2008). This method allowed the reliability performance of power modules to be assessed in real time. A compact thermal model was firstly constructed to investigate the relationship between the power dissipation and the temperature in the power module. Such relationship can be used for fast calculation of junction temperature and the temperatures at each interface inside power modules. The predicted temperature profile was then analyzed using a rainflow counting method so

that the number of thermal cycles with different temperature ranges can be calculated. A reduced order thermo-mechanical model was also constructed to enable a fast calculation of the accumulated plastic strain in the solder material under different loading conditions. The information of plastic strains was then used in the lifetime prediction model to predict the reliability of the solder interconnect under each regular loading condition.

Prognosis with physics-based models are best performed based on an operational profile prediction which must be developed using the steady-state and transient loads, temperatures, or other online measurements. Probabilistic critical-component models then can be ‘‘run into the future’’ by creating statistical simulations of future operating profiles from the statistics of past operational profiles or expected future operating profiles (Vachtsevanos, 2006). Unfortunately, physics of failure models are few and far between, as well as being for more complex processes very difficult and costly to develop. In this case, models are often estimated using system identification techniques, usually via a non-linear least square regression method with a forgetting factor (Ljung, 1999).

Model-based approaches to prognosis differ from feature-based approaches in that they can make RUL estimates in the absence of any measurable events, but when related diagnostic information is present, the model often can be calibrated based on this new information.

A fusion of the feature-based and model-based approaches provides full prognostic ability over the entire life of the component. While failure modes may be unique from component to component, this combined model-based and feature-based methodology can remain consistent across different types of critical components or LRUs.

(Baraldi et al., 2012) propose a prognostic method which predicts the RUL of a degrading system by means of an ensemble of empirical models. RUL predictions of the individual models are aggregated through a Kalman filter (KF)-based algorithm. The method is applied to the prediction of the RUL of turbine blades affected by a developing creep. The Kalman filter assumes that the system must be linear, and the process and measurement noise must be white Gaussian and independent, although non-linearity has been dealt with by the Extended Kalman filter (EFK) and Unscented (UFK).

An Unscented Kalman Filter (UKF) approach is proposed for the purpose of damage tracking and remaining useful life (RUL) prediction of a Polymer electrolyte membrane fuel cells by (Zhang, Xian, and Pierluigi Pisu, 2012). A physics-based, prognostic-oriented catalyst degradation model is developed to characterize the fuel cell damage that establishes the relationship between the operating conditions and the degradation rate of the electro-chemical surface area.

3.6.2 Kalman Filtering

Kalman filtering provides a robust solution that has been successfully applied to a variety of applications like ball gate array solder joints, bearings, batteries, material crack growth, electrolytic capacitors, etc. (Kulkarni, et al., 2012), (Singleton et al., 2015), (Ompusunggu et al., 2015) and (Lall, 2012). Empirical or physics-based degradation models are fused with measurement data within the framework in order to improve the RUL estimation.

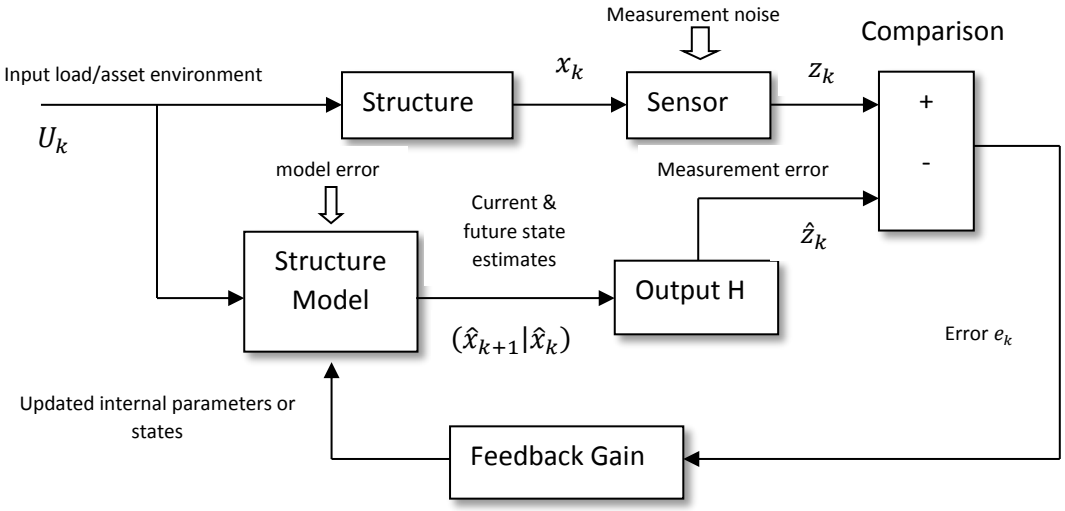


Figure 3.3. Baysian Framework

For linear systems, perturbed by Gaussian white noise, the Kalman filter is an optimal filter. Due to its recursive structure, the filter is also computationally efficient, which is an important consideration in the proposed application. As well as the ability to recursively update model parameters as measurement data becomes progressively available, another attribute of the filter is the ability to estimate model parameters values in the absence of new data.

In the framework proposed in figure (3.3), an ARMA model of the feature vector, replaces the traditional empirical or physics-based degradation models in this Thesis. This is used in conjunction with the Kalman filter and measurement data to enable an estimation of the feature vector to be made up to the failure threshold.

$$\boldsymbol{X}_k = \boldsymbol{A}X_{k-1} + \boldsymbol{B}U_{k-1} + \boldsymbol{G}W_{k-1} \tag{28}$$

$$\boldsymbol{Y}_k = \boldsymbol{H}X_k + \boldsymbol{V}_k \tag{29}$$

The system is corrupted by an additive vector of random noise signals $\boldsymbol{W}_k$, representing all system disturbances, modelling errors etc., and the output is corrupted by another vector of random signals $\boldsymbol{V}_k$, representing measurement noise, discretization errors etc.

The process noise $\boldsymbol{W}_k$ is a vector of zero-mean independent and identically distributed (iid) random variables (RVs), **G** is the stochastic input matrix and $\mathbb{E}\{\boldsymbol{W}_j\boldsymbol{W}_k^T\}$ is the covariance matrix of the stochastic disturbance processes such that:

$$\mathbb{E}\{\boldsymbol{W}_j\boldsymbol{W}_k^T\} = \begin{Bmatrix} \boldsymbol{Q}_W \ \text{for } j = k \\ 0 \text{ for } j \neq k \end{Bmatrix}$$

Where $\mathbb{E}\{.\}$ is the expectation operator and $\boldsymbol{Q}_W$is a positive-semidefinite matrix. The measurement noise $\boldsymbol{V}_k$ is assumed to be composed of Gaussian random variables and its covariance matrix is represented by $\boldsymbol{R} = \mathbb{E}\{\boldsymbol{V}_j\boldsymbol{V}_k^T\}$.The (auto) covariance of a multidimensional random variable Z is expressed as $\mathrm{Cov}(\boldsymbol{Z}) = \mathbb{E}\{(\boldsymbol{Z} - \mathbb{E}\{\boldsymbol{Z}\})(\boldsymbol{Z} - \mathbb{E}\{\boldsymbol{Z}\})^T\}$ but in the case of zero-mean variable it reduces to $\mathrm{Cov}(\boldsymbol{Z}) = \mathbb{E}\{\boldsymbol{Z}_j\boldsymbol{Z}_k^T\}$. The Kalman filter aims at providing

1) $\widehat{\boldsymbol{X}}_k = \mathbb{E}\{\boldsymbol{X}_k|\boldsymbol{X}_{1:k}\}$estimate of the true state mean at any time $t_k > t_0$ given all the previous measurements $\boldsymbol{Y}_i$, *i* = 1,…*k;*
2) $\mathrm{Cov}(\boldsymbol{X}_k - \widehat{\boldsymbol{X}}_k)$ is an estimate of of the state error covariance given all previous measurements.

If the random processes $\boldsymbol{W}_k$ and $\boldsymbol{V}_k$ have a Gaussian distribution, then the conditional state distribution:

$$p(\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}) \sim \mathcal{N}\left(\mathbb{E}\{\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}\}, \mathrm{Cov}(\boldsymbol{X}_k - \widehat{\boldsymbol{X}}_k)\right)$$

The Kalman filter uses two sources of information: (1) a dynamic model expressing the way the system evolves over time and (2) sensory data (measurements). The information from these sources is fused in order to infer an optimal state estimate at every time step.

The Kalman filter algorithm may be broken down into four main stages as shown in the diagram below (Grewal, Mohinder, 2011), (Simon, 2006) & (Kim, 2011).

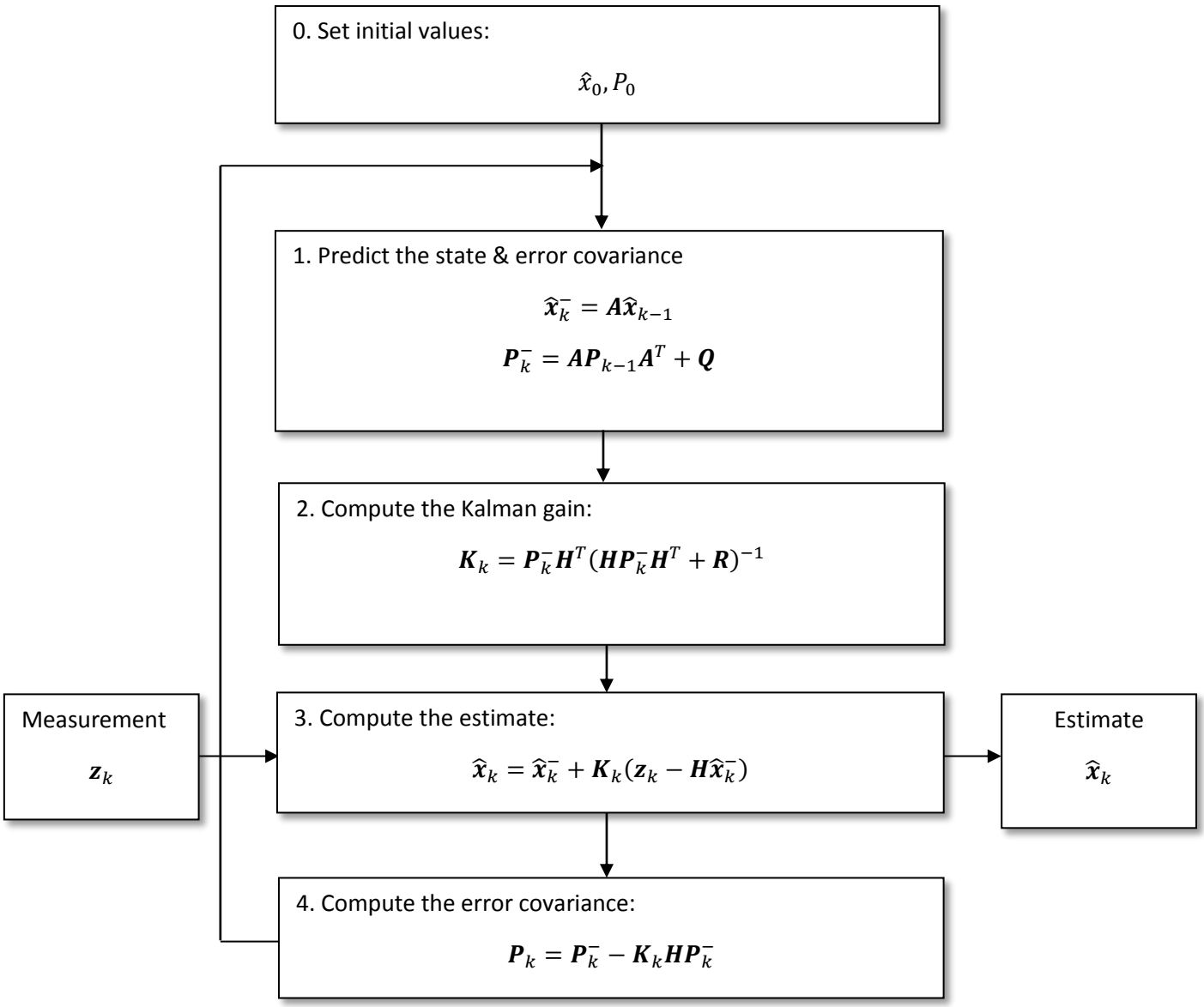


Figure 3.4. Kalman filtering algorithm

It can be seen from the above, that the Kalman filter receives a input in the form of a measurement $\boldsymbol{z_k}$ and returns an estimate of the output $\widehat{\boldsymbol{x}}_k$. The internal process is carried out using four steps. The Kalman filtering algorithm is carried out recursively, hence the subscript k and the subscript '-', means the variable has been predicted.

The first step is for prediction, the two variables $\hat{\boldsymbol{x}}_k^-$ and $\boldsymbol{P}_k^-$, which will be used in the steps 2 through to 4, are computed in this step. It can be seen that the prediction steps $\hat{\boldsymbol{x}}_k^-$ and $\boldsymbol{P}_k^-$ have a very close relationship with the system model **A** and will be discussed more later on.

In step 2, the Kalman gain ($\boldsymbol{K}_k$) is computed. It can be seen that the previously computed variable $\boldsymbol{P}_k^-$ is used. **H** and **R** are values set outside the Kalman filter and are based on process noise and measurement noise defined for the system.

In step 3, an estimate is computed from a measurement given as an input $\boldsymbol{z}_k$, the formula in this step is very closely related to a low-pass filter and the variable $\hat{\boldsymbol{x}}_k^-$ is the one calculated above in step 1.

In step 4, the error covariance is computed. Error covariance is a measure indicating how accurate the estimate is. This forms the basis of the decision making used to trust and use or discard the estimate computed in the previous step.

The Kalman filter algorithm can be summarised into two parts, the prediction process and the estimation process. Where the estimate and error covariance from the previous time point ($\hat{\boldsymbol{x}}_{k-1}$ and $\boldsymbol{P}_{k-1}$, respectively) are used as input and prediction of these two at the current time point ($\hat{\boldsymbol{x}}_k^-$ and $\boldsymbol{P}_k^-$, respectively) are returned as a final result. It is these results that are used for the estimation process. The parameters used in the prediction process ***A*** and ***Q*** are the system model.

3.6.2.1 Estimation Process

The estimation process may be examined in more detail. From the above diagram and description, the estimation process is carried out in steps 2 to 4 and the goal is to compute an estimate, which is the final product of the Kalman filter. In the equation

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}\hat{\boldsymbol{x}}_k^-) \tag{30}$$

first of all a prediction $\hat{\boldsymbol{x}}_k^-$ and a new measurement $\boldsymbol{z}_k$ are required, this is due to the Kalman filter being recursive. The measurement $\boldsymbol{z}_k$is given as an input, for the prediction process, a value is passed from the prediction process (discussed next). **H** is a matrix related to the system model so this is already known. The variable $\boldsymbol{K}_k$

is the Kalman gain, and if the value of this variable is known, a new estimate can be produced. The Kalman gain $\boldsymbol{K}_k$ is computed from

$$\boldsymbol{K}_k = \boldsymbol{P}_k^- \boldsymbol{H}^T (\boldsymbol{H}\boldsymbol{P}_k^- \boldsymbol{H}^T + \boldsymbol{R})^{-1} \quad (31)$$

and forms the weighting process which allows step 3 to perform an estimate. In the Kalman filter, the weighting $\boldsymbol{K}_k$used in the computation of the estimate is calculated every step and it is deemed to be adaptive.

3.6.2.2 Error Covariance

Step 4 in the algorithm, calculates the error covariance.

$$\boldsymbol{P}_k = \boldsymbol{P}_k^- - \boldsymbol{K}_k \boldsymbol{H} \boldsymbol{P}_k^- \quad (32)$$

The error covariance is used to indicate the difference between the estimate from the Kalman filter and the true but unknown value. It gives the degree of accuracy of the estimate. If $\boldsymbol{P}_k$is large, the error of the estimate is large and if $\boldsymbol{P}_k$is small, the error of the estimate is small. The relationship between $\boldsymbol{x}_k$ and its estimate $\hat{\boldsymbol{x}}_k$ and the error covariance is the normal distribution

$$\boldsymbol{x}_k \sim N(\hat{\boldsymbol{x}}_k, \boldsymbol{P}_k) \quad (33)$$

which means the variable $\boldsymbol{x}_k$ follows a normal distribution with mean $\hat{\boldsymbol{x}}_k$ and covariance $\boldsymbol{P}_k$. The Kalman filter computes the probability distribution of the estimate of the variable $\boldsymbol{x}_k$ and selects the highest value probability as an estimate. Error covariance can more be more formal defined mathematically as:

$$\boldsymbol{P}_k = \mathbb{E}\{(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^T\} \quad (34)$$

and is basically the mean square of the error of the estimate, hence the size of the error covariance is proportional to the error of the estimate.

3.6.2.3 Prediction Process

The last step of the Kalman filtering processing is the prediction procedure and relates to step one in the flowchart. The prediction process is concerned with how the estimate $\hat{\boldsymbol{x}}_k$ will vary when time changes from t_k to t_{k+1} is predicted. The equations are repeated below

$$\hat{\boldsymbol{x}}_{k+1}^{-} = \boldsymbol{A}\hat{\boldsymbol{x}}_k \tag{35}$$

$$\boldsymbol{P}_{k+1}^{-} = \boldsymbol{A}\boldsymbol{P}_k\boldsymbol{A}^T + \boldsymbol{Q} \tag{36}$$

where the first equation predicts the estimate and the second equation predicts the error covariance. $\hat{\boldsymbol{x}}_k$ and $\boldsymbol{P}_k$ are the values calculated from steps 3 and 4 respectively, **A** and **Q** are already defined by the system model. The notation used in the subscript, 'k+1' and '-' are the value at time t_{k+1} and that it is a prediction, respectively.

$\hat{\boldsymbol{x}}_k$	Estimate of the state variable
$\hat{\boldsymbol{x}}_k^{-}$	Prediction of the state variable
$\boldsymbol{P}_k$	Estimate of the error covariance
$\boldsymbol{P}_k^{-}$	Prediction of the error covariance

In the Kalman filter the estimate is performed by

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^{-} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}\hat{\boldsymbol{x}}_k^{-}) \tag{37}$$

the estimate from the previous time point ($\hat{\boldsymbol{x}}_{k-1}$) can not been seen in this equation, but the prediction of the current time point ($\hat{\boldsymbol{x}}_k^{-}$) is presented instead. This prediction is obtained by using the estimate from the previous time point. The prediction is calculated from the following

$$\hat{\boldsymbol{x}}_k^{-} = \boldsymbol{A}\hat{\boldsymbol{x}}_{k-1} \tag{38}$$

by substituting this into the expression above, the estimate can be calculated.

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^{-} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}\hat{\boldsymbol{x}}_k^{-})$$
$$\boldsymbol{A}\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}\hat{\boldsymbol{x}}_k^{-}) \tag{39}$$

This enables the estimate to be formed from the previous time step. Sometimes the prediction is referred to as a *priori* estimate and the estimate is called a *posteri*estimate.

3.6.3 Particle Filters

Particle filters were proposed for RUL prediction by (Orchard ME, Vachtsevanos, 2009) and (Saha, Bhaskar, et al. 2009), and have the advantage of not being bound

by the linear system or Gaussian noise assumption. (Zio and Peloni, 2011) proposes a methodology for the estimation of the remaining useful life of components based on particle filtering. The approach employs Monte Carlo simulation of a state dynamic model and a measurement model for estimating the posterior probability density function of the state of a degrading component at future times, in other words for predicting the time evolution of the growing fault or damage state. The proposed approach is applied to a crack fault, with satisfactory results.

(An, Dawn, Joo-Ho, and Nam, 2013) present a Matlab-based tutorial for model-based prognostics, which combines a physical model with observed data to identify model parameters, from which the RUL can be predicted. The particle filter is used in this tutorial for parameter estimation of damage or a degradation model. As examples, a battery degradation model and a crack growth model are used to explain the updating process of model parameters, damage progression, and RUL prediction.

(Daigle and Goebel, 2010) overcame the problem of limited sensor data by applying a model-based prognostics approach using particle filters with application on solenoid valves. (Daigle and Goebel, 2010) also presented a detailed model of the solenoid valve and extended it according to the damage evolving during the valve's lifetime. The measurement models, which establish the mapping from measurement to the internal system state, were the system states themselves, plus measurement noise.

(Lall, Lowe and Goebel, 2010) extended their work on structural damage to Ball Grid Array (BGA) interconnects. In this paper, the effectiveness of the proposed particle filter and resistance spectroscopy based approach in a Prognostic Health Management (PHM) framework has been demonstrated for electronics. The measured state variable has been related to the underlying damage state using non-linear finite element analysis. With the particle filter being used to estimate the state variable, rate of change of the state variable, acceleration of the state variable and construct a feature vector. The estimated state-space parameters have been used to extrapolate the feature vector into the future and predict the time-to-failure at which the feature vector will cross the failure threshold. RUL has been calculated based on the evolution of the state space feature vector.

3.6.3.1. Particle Filter Implementation

The finalised state model needs to be projected until the failure threshold is reached, the particle filter is a powerful and emerging tool for sequential signal processing that is finding application in many engineering and science problems (Arulampalam, Maskell, Gordon, and Clapp, 2002), (Zio and Maio, 2010), (Zio and Peloni, 2011), (An, 2013) and (Saha, 2011) and is used here to determine the RUL due to the ability to add non-linear components to the model if necessary.

The framework below gives a conceptual schematic of a particle-filtering framework for addressing the fault prognosis problem. System sensors and the feature extraction module provide the sequential observation (or measurement) data of the fault growth process z_k at time instant *k*. The fault progression can be explained through the state-evolution model and the measurement mode

$$\boldsymbol{x_k = f_k(x_{k-1}, \omega_k) \leftrightarrow p(x_k|x_{k-1})}$$
$$\boldsymbol{z_k = h_k(x_k, v_k) \leftrightarrow p(z_k|x_k)}$$

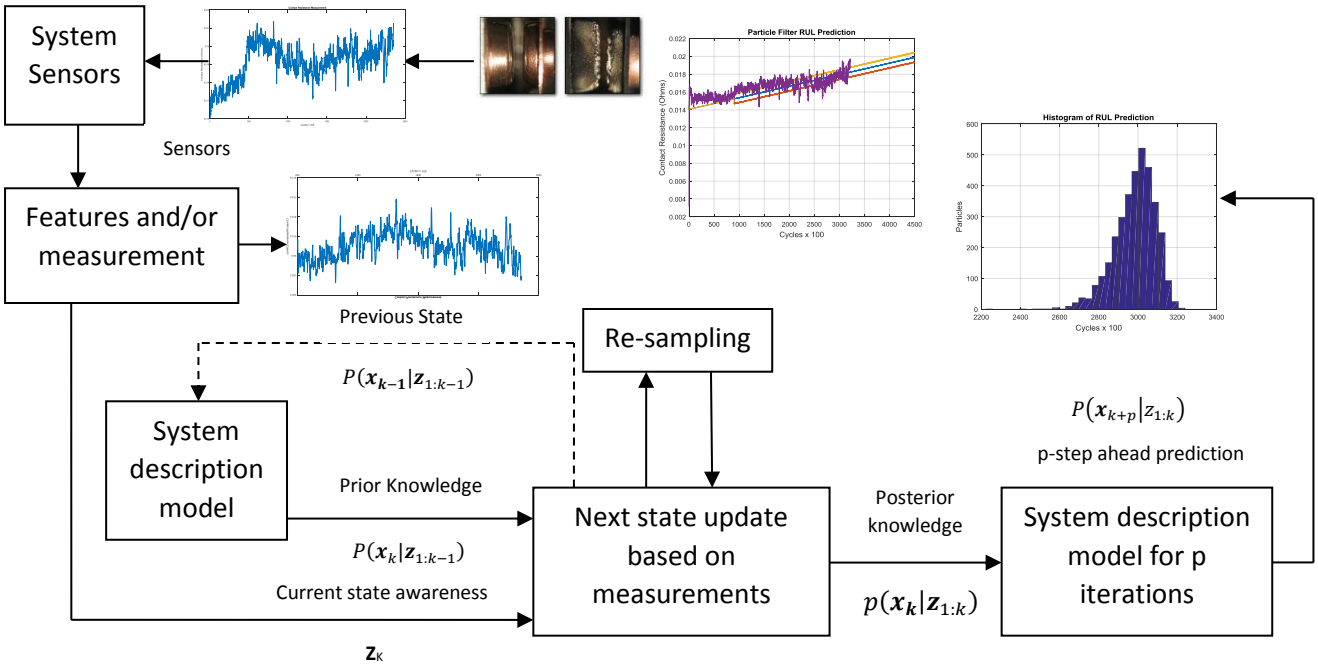


Figure 3.5. Particle filtering schematic

where $\boldsymbol{x_k}$ is the state of the fault dimension (such as the contact resistance), the changing environment parameters that affect fault growth, ω_k and v_k, are the non-Gaussian noises, and f_k and h_k are nonlinear functions.

The object of non-linear filtering is to recursively estimate $\boldsymbol{x}_k$ which in this case includes a set of parameters that define the evolution of the fault condition in time from measurements $\boldsymbol{z}_k \in R^{n_x}$. The measurements are related to the target state via the measurement equation $\boldsymbol{z}_k = h_k(\boldsymbol{x}_k, v_k)$ where h_k is a known, possibly nonlinear function and v_k is a measurement noise sequence. The noise sequences v_k and ω_{k-1} will assumed to be white, with a known probability density function and mutually independent.

The particle filter exists in various forms, however, in terms of usage in literature for prognostics, the Sampling Importance Re-sampling (SIR) concept is the most prominent. The SIR algorithm proposes a posterior filtering distribution denoted as $\pi(x) = p(\boldsymbol{x}_k|\boldsymbol{z}_k)$ is approximated by a set of N weighted particles $\{\langle \boldsymbol{x}_k^i, w_k^i \rangle; i = 1, \dots, N\}$ sampled from an arbitrarily proposed distribution q(x) that is somewhat similar to $\pi(x)$ (e.g., $\pi(x) > 0 \rightarrow q(x) > 0$ for all $\boldsymbol{x} \in R^{n_x}$. The importance weights w_k^i are proportional to the likelihood $\mathrm{p}(\boldsymbol{z}_k|\boldsymbol{x}_k^i)$ associated to the sample $\boldsymbol{x}_k^i$ and this is then normalised to give the following

$$w_k^i = \frac{\pi(\boldsymbol{x}_k^i)/q(\boldsymbol{x}_k^i)}{\sum_{j=1}^{N} \pi(\boldsymbol{x}_k^j)/q(\boldsymbol{x}_k^j)} \tag{40}$$

such that $\sum_{j=1}^{N} w_k^i = 1$ and the posterior distribution (which is essentially the target distribution) can be approximated as

$$p(\boldsymbol{x}_k|\boldsymbol{z}_k) = \sum_{j=1}^{N} w_k^i \, \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i) \tag{41}$$

As with any process that uses a Bayesian update, the filtering step consists of two stages, firstly the production of a priori state density estimate which is the prediction step and secondly a update of the estimation according to the new measurement information.

The prediction step is given by

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{k-1}) \approx \sum_{i=1}^{N} w_{k-1}^i \delta(\boldsymbol{x}_k - f_{k-1}(\boldsymbol{x}_{k-1}^i) - \omega_{k-1}^i) \tag{42}$$

and the update step is used to modify the particle weights according to the relation

$$\overline{w}_k^i = w_{k-1}^i \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k^i) p(\boldsymbol{x}_k^i|\boldsymbol{x}_{k-1}^i)}{q(\boldsymbol{x}_k^i|\boldsymbol{x}_{k-1}^i, \boldsymbol{z}_k)}$$

$$w_k^i = \frac{\bar{w}_k^i}{\sum_{j=1}^{N} \bar{w}_k^j} \tag{43}$$

Re-sampling may be used if the weights degenerate such that they are close to zero, as this can occur if the system state is poorly represented, it also wastes computing resources on superfluous calculations.

To enable the particle filter to be used for prognosis, a particle filtering based prognostics approach needs to project the current PDF estimate of the damage state in time. The simplest implementation that can be used to solve this problem uses the damage state equation recursively to propagate the posterior PDF estimate defined by $\{\langle \boldsymbol{x}_p^i, w_p^i \rangle; i = 1, \dots, N\}$ in time until $\boldsymbol{x}_p^i$ fails to no longer meet the system failure criteria at some time t_{EOL}^i. Therefore, the RUL PDF is the distribution of $p(t_{EOL}^i - t_p)$ given by the distribution of w_p^i.

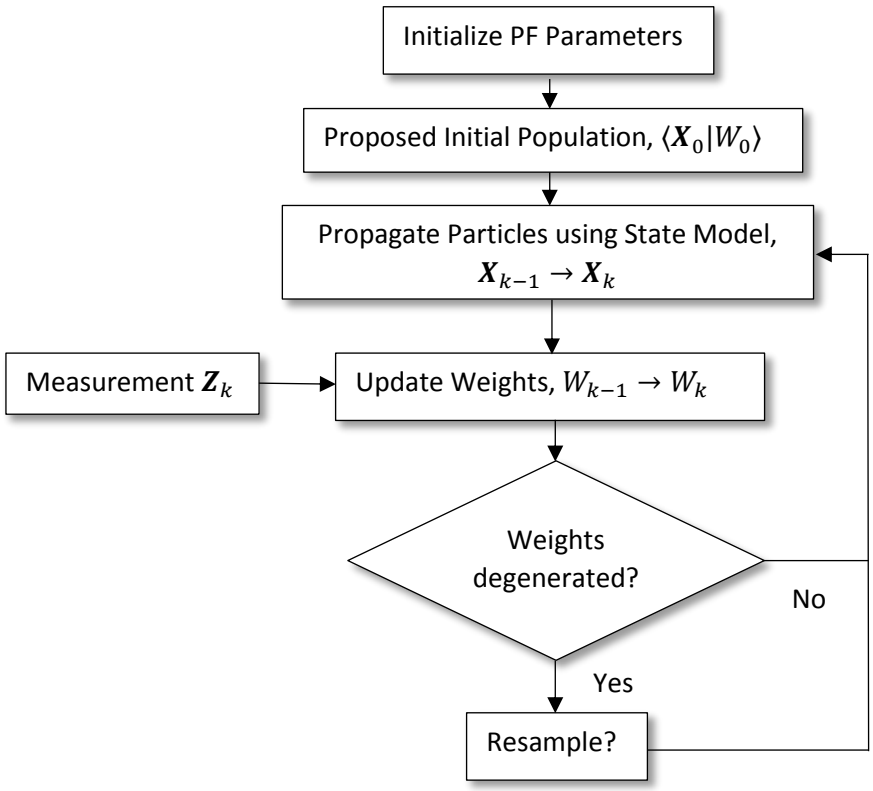


Figure 3.6. Particle filtering flowchart.

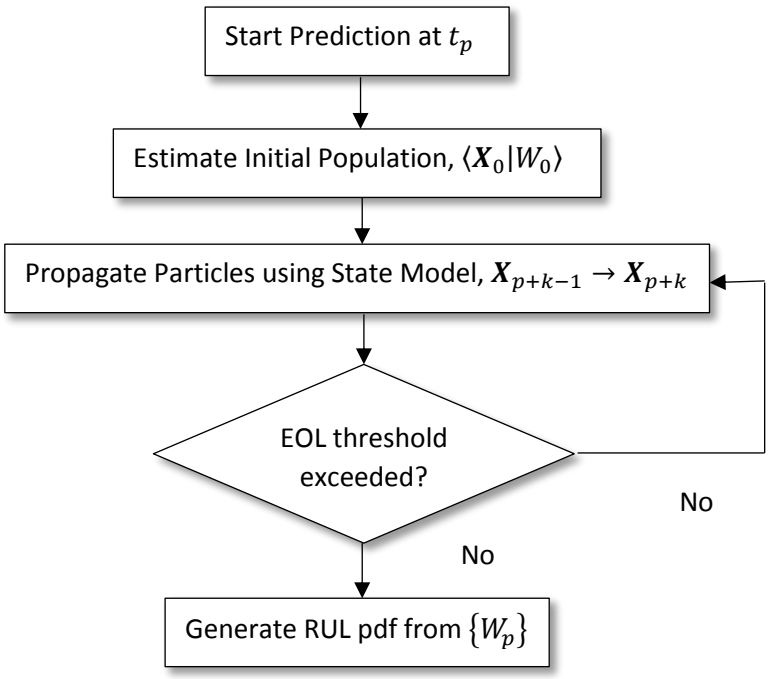


Fig 3.7. Particle filter for RUL prediction flowchart (Saha, 2011)

3.7 Data Driven Techniques

Data-driven models rely only on previously observed data to predict the projection of a system's state or to match similar patterns in the history to infer RUL (Liao and Kottig, 2014). In a lot of cases, historical fault/failure data in terms of time plots of various signals leading up to failure, or statistical data sets may be available.

The derivation of any sort of prediction model may, however be difficult to obtain. In such situations, the use of nonlinear network approximations that are based upon using well-established formal algorithms can be used to provide desired outputs directly in terms of the data. Non-linear networks include the concept of artificial neural networks, fuzzy-logic systems and other computation intelligence methods which are based on the linguistic and reasoning abilities of humans. The techniques provide structured nonlinear function mapping with very desirable properties between the available data and the desired outputs (Vachtsevanos, 2006).

In prediction, Artificial Neural Networks (ANNs), Fuzzy Systems, and other computational intelligence methods have provided an alternative tool for both forecasting researchers and practitioners (Sharda, Ramesh, 1994).

3.7.1 Statistical Models

(Si et al., 2011) produced a comprehensive study on statistical data-driven models for RUL estimation. Two categories of statistical model were identified based upon the CBM data used. The two identified categories of CBM data are firstly; direct CBM data, which indicates the health level of the system directly (e.g. wear level, crack length) and secondly indirect CBM data, which reflects the underlying state of the system health (e.g. contact resistance, temperature, pressure, moisture, humidity, loading, speed and environmental data).

Under the first category, techniques that have been utilised are Regression-based models, Gamma and Weiner based processes and Markovian-based models. Whereas the second category consists of stochastic filtering-based models, covariate-based hazard models and Hidden Markov Model (HMM) and Hidden Semi-Markov Models (HSMM).

Brownian motion (or Wiener Processes) use a Probability density function (PDF) which is an inverse Gaussian function to produce a continuous state data driven prognostic. By using only the current health state PDF the process does not need posterior data. (Wang and Carr, 2010) used a stochastic degradation model based upon an improved version of Brownian motion to predict the RUL of monitored plants. Using a Kalman filter the parameters where drifted in the Brownian motion model up to a failure distribution threshold as opposed to a constant threshold.

In some case the degradation processes are monotonic and evolving only in one direction, such as wear processes or fatigue crack propagation for example. A Gamma process is a natural model for the degradation processes in which the deterioration is supposed to take place gradually over time in a sequence of tiny positive increments (Si et al., 2011). A Gamma process was used by (Lawless and Crowder, 2004) in determining the optimal inspection and maintenance decisions for crack growth based data. (Kuniewslci et al. 2009) combined a Gamma process with a Poisson process for modelling the degradation and the initial defect in order to determine the lifetime distribution.

The underlying assumptions of Markovian-based models are twofold. One is that the future degradation state of the item depends only on the current degradation state,

which is often termed as being memoryless. The other is that the system's state can be revealed directly by the observed CBM information, (Si et al., 2011).

(Zhang, et al., 2005) presented an integrated fault diagnostic and prognostic approach for bearing health monitoring and condition-based maintenance. The proposed scheme consisted of three main components, Principal Component Analysis (PCA), Hidden Markov Model (HMM) and an adaptive stochastic fault prediction model. The principal signal features extracted by PCA are utilized by HMM to generate a component health/degradation index, which is the input to an adaptive prognostics component for on-line remaining useful life prediction.

Fault classification and prognostics applications were proposed by (Dong et al., 2006), (Dong and He, 2007b), (Dong and He, 2007a) using several Hidden Semi-Markov Model (HSMM) for the Blackhawk UH-60A main transmission. Within the planetary carriers, the HSMMs are used to generate observations and estimate the maintenance durations from the acquired training data. This differs from the use of HMMs and can only produce a single observation for each state in turn.

(Alireza, et al. 2015) presented a computationally efficient prognostic methodology for an isolated-gate bipolar transistor (IGBT). The proposed approach is developed by analysing failure mechanisms and statistics of IGBT degradation data obtained from an accelerated aging experiment. The approach explores various probability distributions for modelling discrete degradation profiles of the IGBT component. This allows the stochastic degradation model to be efficiently simulated, in this particular example~1000 times more efficiently than Markov approaches.

Statistical reliability or usage-based prognostic approach may be a viable approach in situations where a viable prognostic model is unavailable or cannot be utilised (this may be due to cost, low failure rate or perhaps the poor availability of sensor networks within the equipment). Of all the prognostic algorithm approaches this is least complex and requires only that a history of component failure or operational usage profile data is available. One typical approach would be to fit a Weibull distribution (or other statistical failure distribution) to such failure or inspection data (Groer, 2000) and (Schomig and Rose, 2003).

This approach obviously suffers from a lack of information that is present in model and data driven techniques (condition based techniques), but can still be used to good effect to drive an interval-based maintenance schedule (several commercial packages

are available that use this type of approach). The update being performed by revising the information obtained from maintenance and hence this technique is reliant upon a regular updated maintenance database.

Weibull distributions are a class of reliability functions that have been used for data-driven prognostics. A three-parameter Weibull failure rate function for wind turbine reliability assessment was presented by (Guo, Watson, Tavner, and Xiang, 2009). The proposed model dealt better with incomplete field failure data than the traditional Weibull model, which in fact is a special case of the introduced three-parameter model. An approach for deriving and using accurate Weibull distributions was presented by (Sutherland et al. 2003) for the means of identification of subpopulations of related systems in the maintenance database for condition-based maintenance of motors.

A Weibull distribution was used to model the time-to failure for the RUL prediction using a statistical model for an application of pumps in a hot strip steel mill by (Goode et al., 2000).

Combining vibration data with reliability data, RUL was calculated during the potential failure to functional failure interval.

The reliability functions, which require data associated with individual failure mode(s), can be developed with sufficient historical data. The method incorporates system age and independent hazards into RUL prediction. It is easy to build a statistically adequate model; however, the model may not follow the underlying degradation process. The parameters are often selected by experts, and that process is time consuming. The reliability function may require assumptions such as the times to failure must be independent and identical distributed, all covariates that affect the hazard rate must be included in the model, and the effect of covariates is statistically time dependent (Liao and Kottig, 2014).

Autoregressive moving average (ARMA), Autoregressive integrated moving average (ARIMA) and ARMAX models are widely used for modelling and forecasting time series data (Box and Jenkins, 1976). The future value is assumed to be a linear function of past observations and random errors and each of the three types of autoregressive models vary slightly in the linear equation used to relate inputs, outputs and noise. ARMA and ARMAX models can remove temporal trends, they should only be used for stationary data. A time series is designated to be (weakly) stationary if its first two moments (mean and variance) are time-invariant under translation (Tsay, 2000). The autocorrelation also needs to be independent of time. Consequently, a

trend test should be performed prior to modelling to verify whether a stationarity assumption is appropriate. ARIMA models overcome this weakness by enforcing integration and thus are useful in describing systems with low frequency disturbances (i.e. trends), (Sikorska, Hodkiewicz and Ma, 2011).

Autoregressive models are developed in three recursive steps, (Sikorska, Hodkiewicz and Ma, 2011):

(i) Model identification: Using a set of time series data, values for the orders of the autoregressive and moving average parts of the ARMA/ARIMA equations are proposed, as well as the regular-difference parts for the ARIMA model.

(ii) Parameter estimation: Using non-linear optimization techniques (e.g. least-squares method), parameters of the ARMA/ARIMA equations are calculated to minimise the overall error between the model output and observed input-output data.

(iii) Model validation: A number of standard diagnostic checks can be employed to verify the adequacy of ARMA/ARMAX models. (Wu, 2007) options include the following: examining standardized residuals, autocorrelation of residuals, final prediction error (FPE), Akaike Information criterion (AIC), and Bayesian Information Criterion (BIC).

(Lee, Jay, et al., 2006) predicted the behaviour of machining process spindle load signatures using ARMA modelling techniques. Load sensor readings from a boring machine spindle were remotely collected and processed into joint time-frequency (TF) distributions. Performance related signatures were extracted from the TF distributions using the TF moments and Principal Component Analysis. ARMA modelling techniques were then utilized to predict the behaviour of the extracted principal components.

(Guclu, Adem, et al., 2010) Presented an ARMA based prognostics solution for railway turnouts. The model is applied to data collected from real turnout systems and the failure progression is obtained manually using the exponential degradation model. RUL of ten turnout systems were reported and results looked promising.

(Saha, 2009) made a comparison of prognostic algorithms for estimating RUL of batteries. An ARIMA model was compared to other models such as Particle filtering and it was concluded that being a purely data-driven method, does not incorporate

any physics of the process into the computation, and hence ends up with wide uncertainty margins that make it unsuitable for long term predictions.

3.7.2 Damage Modelling With ARMA Series

The ability to build the state space model from prior measurement data as the system degrades allows complex processes such as relay contacts to be modelled more accurately and incorporates adaptation to the uncertainties described in chapter 2.

A dynamic system with stochastic properties described in the discrete linear form as

$$\boldsymbol{X}_k = \boldsymbol{A}X_{k-1} + \boldsymbol{B}U_{k-1} + \boldsymbol{G}W_{k-1} \tag{44}$$

$$\boldsymbol{Y}_k = \boldsymbol{H}X_k + \boldsymbol{V}_k \tag{45}$$

is called the state transition equation. This equation determines how each element of the state vector evolves over time. In the discrete linear case each element of the state vector at time index k+1 is expressed as a linear combination of all the elements of the state vector at time k plus some linear combination of the elements of an input vector U_k where k represents the time index. The matrix **A** is called state transition matrix and the matrix **B** is called input matrix. Because in general the states themselves are not directly measurable, Eq. (45) indicates how the measurements are linked to the system states. More precisely this equation expresses each element of the measurement vector **Y** as a function of the state vector elements plus some input. This is a static equation. The matrix **H** is called measurement (or output) matrix.

The relationship between the input and output is given by

$$Y(z) = H(z\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B}U(z) = H(z)U(z) \tag{46}$$

i.e. where *H(z)* is the transfer function with numerator and denominator are represented by polynomials

$$H(z) = \frac{z^d \sum_{l=0}^{m} b_i z^{m-l}}{z^n + \sum_{i=1}^{n} a_i z^{n-i}} z^d \tag{47}$$

where the denominator degree is *n*, the numerator degree is $m \leq n$, and the relative degree is $n - m$. This can be written in autoregressive moving-average (ARMA) form as

$$y_k = -a_1 y_{k-1} - \cdots a_n y_{k-n} + b_0 u_{k-d} + b_1 u_{k-d-1} + \cdots + b_m u_{k-d-m}$$

$$y_k = -\sum_{i=1}^{n} a_i y_{k-i} + \sum_{i=0}^{m} b_i \, u_{k-i-d} \qquad (48)$$

The system delay is d = n-m.

In order to determine the system degree the input to the system, U_k must excite all modes of the unknown system. It can be shown that one can solve for the n+m+1 unknown ARMA parameters if U_k is Persistently Exciting (PE) of order n+m+1 (Marafioti et al., 2014).

For the relay, a square wave representing the switching input was used as the excitation signal U_k and a (2,1) ARMA series was found to be adequate to model the contact degradation.

To determine the unknown parameters, the general ARMA form is written as

$$y_k = [-y_{k-1} - y_{k-2} \ldots - y_{k-n} \, u_{k-d} \ldots \, u_{k-d-m}] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ b_0 \\ \vdots \\ b_m \end{bmatrix} + v_k$$

or $$y_k = \boldsymbol{h}_k^T \boldsymbol{\theta} + v_k \qquad (49)$$

where the unknown system parameter vector is

$$\boldsymbol{\theta} = [a_1 \cdots a_n \; b_0 \; \cdots b_m]^T \qquad (50)$$

and the known regression matrix is given in terms of previous outputs and inputs by

$$\boldsymbol{h}_k^T = [-y_{k-1} - y_{k-2} \ldots - y_{k-n} \, u_{k-d} \ldots \, u_{k-d-m}] \qquad (51)$$

and v_k is the measurement noise.

By using regression based methods the parameter of $\boldsymbol{\theta}$ may be estimated. Since there are $n + m$ +1 parameters to estimate, one needs n previous output values and m +1 previous input values.

If $y(n)$ and $h_1(n), h_2(n) \ldots h_k(n)$ are the observation sequences of previous outputs and inputs of *y* and *h* at $t_1, t_2 \ldots t_m$ then the relationship between *y* and *h* and the unknown system parameters can be expressed as

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{bmatrix} = \begin{bmatrix} h_1(1) & \cdots & h_k(1) \\ h_1(2) & \cdots & h_k(2) \\ & \vdots & \\ h_1(n) & \cdots & h_k(n) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}, \text{ or } \boldsymbol{y} = \boldsymbol{H\theta}$$

gives the estimate

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{H}^{-1}\boldsymbol{y} \tag{52}$$

where $\widehat{\boldsymbol{\theta}}$ is the parameter estimation and $\boldsymbol{H}^{-1}$is the inverse of the **H** matrix.

$$\boldsymbol{\varepsilon} = y - \boldsymbol{H\theta} \tag{53}$$

Where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2 \cdots \varepsilon_m)^T$ is the error vector. Hence for a cost function

$$\boldsymbol{J} = \sum_{i=1}^{m} \varepsilon_i^2 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} \tag{54}$$

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T y \tag{55}$$

The Recursive Least Square Estimation (RLS) is one such regression based method that may be used for the parameter estimation (Ljung, 1999.). In the following equation one must know the system degree and delay in order to construct the regression vector h_n in equation (51)**.**

The general form of a recursive algorithm is

$$\widehat{\boldsymbol{\theta}}(n) = \widehat{\boldsymbol{\theta}}(n-1) + \boldsymbol{K}(n)\boldsymbol{\varepsilon}(n) \tag{56}$$

Using n-1 groups of data the least square estimation is show as follows:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{H}_{n-1}^T\boldsymbol{H}_{n-1})^{-1}\boldsymbol{H}_{n-1}^T\boldsymbol{Y}_{n-1} \tag{57}$$

The data at n is given by

$$\boldsymbol{Y}_n = \boldsymbol{H}_n\boldsymbol{\theta}(n) \tag{58}$$

where

$$\boldsymbol{Y}_n \begin{bmatrix} y(1) \\ \vdots \\ y(n-1) \\ \cdots \\ y(n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_m \\ \cdots \\ y(n) \end{bmatrix} \tag{59}$$

$$\boldsymbol{H}_n = \begin{bmatrix} h_1(1) & \cdots & h_k(1) \\ \vdots & \cdots & \vdots \\ h_1(n-1) & \cdots & h_k(n-1) \\ \cdots & \cdots & \cdots \\ h_1(n) & \cdots & h_k(n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_n \\ \cdots \\ h^T(n) \end{bmatrix} \tag{60}$$

The new least square estimation is given as follows:

$$\widehat{\boldsymbol{\theta}}(n) = (\boldsymbol{H}_n^T \boldsymbol{H}_n)^{-1} \boldsymbol{H}_n^T \boldsymbol{Y}_n \tag{61}$$

The recursive solution of this equation is well documented (Haykin, 2013)

$$\boldsymbol{\varepsilon}(n) = y(n) - h^T(n)\widehat{\boldsymbol{\theta}}(n-1) \tag{62a}$$

$$\boldsymbol{P}(n) = (\boldsymbol{P}^{-1}(n-1) + h(n)h^T(n))^{-1} \tag{62b}$$

$$\boldsymbol{K}(n) = \boldsymbol{P}(n)h(n) \tag{62c}$$

$$\widehat{\boldsymbol{\theta}}(n) = \widehat{\boldsymbol{\theta}}(n-1) + \boldsymbol{K}(n)\varepsilon(n) \tag{62d}$$

However a matrix inverse is still required in equation (62b), this can be avoided by employing the matrix lemma.

If **A**, **C**, **BCD** are non-singular square matrices, then

$$[\boldsymbol{A} + \boldsymbol{BCD}]^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}[\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B}]^{-1}\boldsymbol{DA}^{-1} \tag{63}$$

With $\boldsymbol{A} = \boldsymbol{P}^{-1}(n-1)$, $\boldsymbol{B} = h(n)$, **C**=1, and $\boldsymbol{D} = h^T(n)$

$$\boldsymbol{P}(n) = (\boldsymbol{P}^{-1}(n-1) + h(n)h^T(n))^{-1}$$

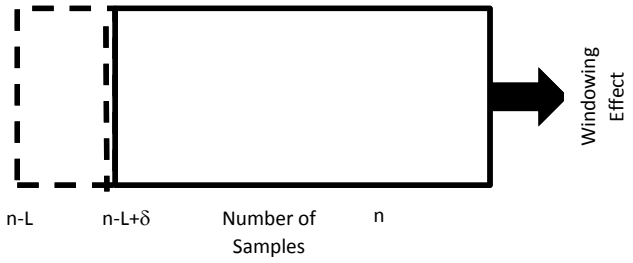
$$\boldsymbol{P}(n) = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}[\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B}]^{-1}\boldsymbol{DA}^{-1}$$

$$\boldsymbol{P}(n) = \boldsymbol{P}(n-1) - \frac{\boldsymbol{P}(n-1)h(n)h^T(n)\boldsymbol{P}(n-1)}{[1+h^T(n)\boldsymbol{P}(n-1)h(n)]} \tag{64}$$

Where equation (62b) can be replaced with equation (64) a recursive update of the covariance $\boldsymbol{P}(n)$.

3.7.2.1 Sliding Window Recursive Least Square ARMA Parameter Estimation

The RLS algorithm has been used for system identification and time-series analysis (Ljung, 1999). However, the RLS algorithm has infinite memory, in the sense that all the data from time n=0, will affect the coefficients of y_k, thus becoming computationally inefficient. For a non-stationary process, whose statistics are changing in time, several options exist, firstly, is the use of an exponentially weighted factor which gradually reduces the importance of past data (Haykin, 2013), (Astrom and Wittenmark, 1995). The second option is to use to minimise the sum of squares error over a finite window. The sliding window RLS algorithm benefits from only an incremental amount of computation being needed at each time step in which the estimate is a function of data over a rectangular window of length n: $\{y(n-L+\delta) \ldots, y(n)\}$. Where δ is the step size (Mulgrew and Cowan, 2012), (Zhao et al. 1994), (Jiang and Zhang, 2004), (Belge and Miller, 2000). From the diagram below, the procedure is to choose an observation period of *L* samples which, is accurate enough give an estimate of desired accuracy. Typically this may involve looking at the error covariance matrix obtained for n samples $P_n = \sigma^2 P_n$ is satisfied to a degree of acceptability.



The sliding window algorithm consists of two distinct steps. Firstly, a new data point y(L+δ) is added. Then an old data point y(n-L+δ) is discarded, thus keeping the number of active points equal to n.

Letting $\boldsymbol{P}(n) = (\boldsymbol{H}_n^T \boldsymbol{H}_n)^{-1}$, $\widehat{\boldsymbol{\theta}}(n) = \boldsymbol{P}(n)\boldsymbol{H}_n^T \boldsymbol{Y}_n$ for the parameter estimation given by $(h(k), y(k), k = 1,2, \ldots, n$

Where

$$\boldsymbol{H}_n = \begin{bmatrix} h^T(1) \\ h^T(2) \\ \vdots \\ h^T(n-1) \end{bmatrix}, \ \boldsymbol{Y}_n = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n-1) \end{bmatrix} \quad (65)$$

If the data group is increased to $h(n), y(n)$ then the parameter estimation is given by

$$\overline{\boldsymbol{\theta}}(n) = (\boldsymbol{H}_n^T \boldsymbol{H}_n)^{-1} \boldsymbol{H}_n^T \boldsymbol{Y}_n$$

Using the matrix inversion lemma (63).

$$\overline{\boldsymbol{\theta}}(n) = \widehat{\boldsymbol{\theta}}(n-1) + \frac{\boldsymbol{P}(n-1)h(n)\left[y(n) - h^T(n)\widehat{\boldsymbol{\theta}}(n-1)\right]}{1 + h^T(n)\boldsymbol{P}(n-1)h(n)} \quad (66)$$

Again, using the matrix inversion lemma (63).

$$\overline{\boldsymbol{P}}(n) = (\boldsymbol{H}_n^T \boldsymbol{H}_n)^{-1}$$

$$\overline{\boldsymbol{P}}(n) = \boldsymbol{P}(n-1) - \frac{\boldsymbol{P}(n-1)h(n)h^T(n)\boldsymbol{P}(n-1)}{1 + h^T(n)\boldsymbol{P}(n-1)h(n)} \quad (67)$$

Where

$$\boldsymbol{H}_n = \begin{bmatrix} h^T(1) \\ h^T(2) \\ \vdots \\ h^T(n-1) \\ h^T(n) \end{bmatrix} \qquad \boldsymbol{Y}_n = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n-1) \\ y(n) \end{bmatrix} \quad (68)$$

Therefore, $\boldsymbol{H}_n = \begin{bmatrix} h^T(1) \\ \boldsymbol{H} \end{bmatrix}$

If the first group of data $h(n),\ y(n)$ is eliminated and the parameter estimation becomes

$$\boldsymbol{P}(n) = (\boldsymbol{H}^T \boldsymbol{H})^{-1}, \boldsymbol{H} = \begin{bmatrix} h^T(2) \\ \vdots \\ h^T(n-1) \\ h^T(n) \end{bmatrix},$$

$$\boldsymbol{H}_n^T \boldsymbol{H}_n = \boldsymbol{H}^T \boldsymbol{H} + h(1)h^T(1),$$

$$\boldsymbol{P}(n) = (\boldsymbol{H}_n^T \boldsymbol{H}_n - h(1)h^T(1))^{-1}$$

$$\boldsymbol{P}(n) = ((\overline{\boldsymbol{P}}(n))^{-1} - h(1)h^T(1))^{-1} \quad (69)$$

Again employing the matrix lemma (63)

$$\boldsymbol{P}(n) = \overline{\boldsymbol{P}}(n) - \frac{\overline{\boldsymbol{P}}(n)(-h(1))h^T(1)\overline{\boldsymbol{P}}(n)}{1 + h^T(1)\overline{\boldsymbol{P}}(n)\left(-h(1)\right)}$$

$$\boldsymbol{P}(n) = \overline{\boldsymbol{P}}(n) + \frac{\overline{\boldsymbol{P}}(n)h(1)h^T(1)\overline{\boldsymbol{P}}(n)}{1-h^T(1)h(1)\overline{\boldsymbol{P}}(n)} \tag{70}$$

The parameter estimation is therefore given by

$$\widehat{\boldsymbol{\theta}}(n) = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{Y} = \boldsymbol{P}(n)\boldsymbol{H}^T\boldsymbol{Y} \tag{71}$$

Where

$$\big(h(k), y(k)\big), \qquad k = 2, \dots, n,$$

$$\boldsymbol{Y} = \begin{bmatrix} y(2) \\ \vdots \\ y(n-1) \\ y(n) \end{bmatrix} \tag{72}$$

$$\boldsymbol{H}_n^T\boldsymbol{Y}_n = \boldsymbol{H}^T\boldsymbol{Y} + y(1)h(1) \tag{73a}$$

$$\widehat{\boldsymbol{\theta}}(n) = \boldsymbol{P}(n)\left(\boldsymbol{H}_n^T\boldsymbol{Y}_n - y(1)h(1)\right) \tag{73b}$$

$$= \overline{\boldsymbol{P}}(n) + \frac{\overline{\boldsymbol{P}}(n)h(1)h^T(1)\overline{\boldsymbol{P}}(n)}{1-h^T(1)h(1)\overline{\boldsymbol{P}}(n)}$$

$$= \overline{\boldsymbol{P}}(n)\boldsymbol{H}_n^T\boldsymbol{Y}_n - \frac{y(1)\overline{\boldsymbol{P}}(n)h(1)h^T(1)\overline{\boldsymbol{P}}(n)h(1)}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)} - y(1)\overline{\boldsymbol{P}}(n)h(1) + \frac{\overline{\boldsymbol{P}}(n)h(1)h^T(1)\overline{\boldsymbol{P}}(n)\boldsymbol{H}_n^T\boldsymbol{Y}_n}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)}$$

$$= \overline{\boldsymbol{\theta}}(n) - \frac{y(1)\overline{\boldsymbol{P}}(n)h(1)}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)} \times \left(h^T(1)\overline{\boldsymbol{P}}(n)h(1) + 1 - h^T(1)\overline{\boldsymbol{P}}(n)h(1)\right)$$

$$+ \frac{\overline{\boldsymbol{P}}(n)h(1)}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)} \times h^T(1)\overline{\boldsymbol{\theta}}(n$$

$$= \overline{\boldsymbol{\theta}}(n) - \frac{y(1)\overline{\boldsymbol{P}}(n)h(1)}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)} + \frac{\overline{\boldsymbol{P}}(n)h(1)h^T(1)\overline{\boldsymbol{\theta}}(n)}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)}$$

$$\widehat{\boldsymbol{\theta}}(n) = \overline{\boldsymbol{\theta}}(n) + \frac{\overline{\boldsymbol{P}}(n)h(1)\left(h^T(1)\overline{\boldsymbol{\theta}}(n)-y(1)\right)}{1-h^T(1)\overline{\boldsymbol{P}}(n)h(1)} \tag{74}$$

Therefore, at any point n, window length L and δ=1

$$\widehat{\boldsymbol{\theta}}(n) = \overline{\boldsymbol{\theta}}(n) + \frac{\overline{\boldsymbol{P}}(n)h(n-L+1)h^T(n-L+1)\overline{\boldsymbol{\theta}}(n)-y(n-L+1)}{1-h^T(n-L+1)\overline{\boldsymbol{P}}(n)h(n-L+1)} \tag{75}$$

$$\boldsymbol{P}(n) = \overline{\boldsymbol{P}}(n) + \frac{\overline{\boldsymbol{P}}(n)h(n-L+1)\overline{\boldsymbol{P}}(n)h^T(n-L+1)}{1-h^T(n-L+1)\overline{\boldsymbol{P}}(n)h(n-L+1)} \tag{76}$$

The parameter estimation is related to the current n data samples, with the past data having no effect. The computation is double that of the standard RLS algorithm, but still allows efficient speeds of calculation to be obtained and limits the computation in terms of data size, all of which is desirable when dealing with large data sets and limited on board processing. The estimated parameters within the sliding window $\widehat{\boldsymbol{\theta}}(n-L+1)$ are then used to form the state space model.

$$x_{k+1} = \boldsymbol{A}x_k + \boldsymbol{B}u_k \tag{77}$$

$$\boldsymbol{x}_{k+1} = \begin{bmatrix} \boldsymbol{x}_{1(k+1)} \\ \boldsymbol{x}_{2(k+1)} \\ \vdots \\ \boldsymbol{x}_{n(k+1)} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ a_0 & a_1 & a_2 & a_3 & \cdots & a_{n-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{1(k)} \\ \boldsymbol{x}_{2(k)} \\ \vdots \\ \vdots \\ \boldsymbol{x}_{n(k)} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ b_0 \end{bmatrix} u_k \tag{78}$$

The SWRLS ARMA models fit to the Median filtered data is shown below in figure 3.8, as is evident, a (2,1) ARMA model forms a good fit to the data.

The magnitude of error squared between the estimated and actual data in the window may be calculated, as shown in figure 3.9. The original and the estimated data set are overlaid with an error magnitude of the order of 10^{-9}. The accuracy of the model is important in the next stage, where Kalman filtering is used to extrapolate the current measurement and the model at time n, up to the threshold of failure.

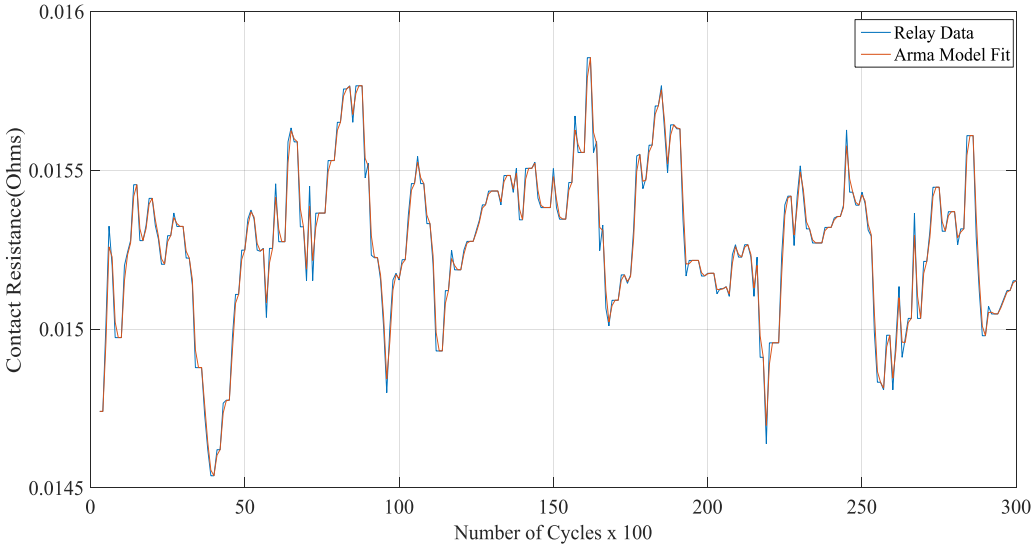


Figure 3.8. ARMA model fit to a section of median filtered data

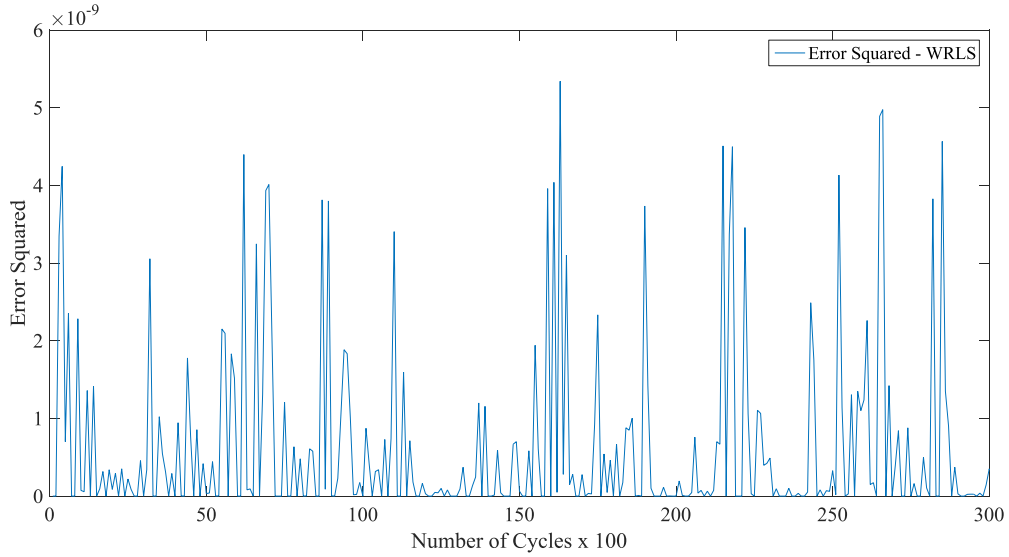


Figure 3.9. ARMA model fit squared error

3.7.2 Artificial Intelligence-Based Models

(Werbos, 1988) first reported that Artificial Neural Networks (ANNs) trained with the back propagation algorithm outperform traditional statistical methods such as regression and Box-Jenkins approaches. Pursuing literature, ANNs are maybe the most popular machine learning based methodology used for prognostics. The idea of ANNs is to mimic the neural structure of the human brain by using input, hidden and output layers that interact with one another via weighted connections. They may be divided into two main classes; namely, those which use supervised learning and those

which use unsupervised learning, in both cases, they may be thought of non-linear black box function approximators, with multiple input/outputs. Condition monitoring data is used with target data e.g. health states/indicators in order to train the network in supervised learning. Whereas the unsupervised approach employs learning models in the form of algorithms (e.g. SOM and K-means) to find hidden structure in the data. Unlike the traditional model-based methods, ANNs are data-driven and self-adaptive, and they make very few assumptions about the models for problems under study. ANNs learn from examples and attempt to capture the subtle functional relationship among the data. Thus, ANNs are well suited for practical problems, where it is easier to have data than knowledge governing the underlying system being studied. Generally, they can be viewed as one of many multivariate nonlinear and nonparametric statistical methods (Cheng, Bing, and Titterington, 1994).

The main problem of ANNs is that their decisions are not always evident and the training time can be somewhat dependent upon the quantity of the data. Nevertheless, they provide a feasible tool for practical prediction problems. Data-driven approaches to failure prognosis also take advantage of recurrent neural networks, dynamic wavelet neural networks, neuro-fuzzy systems, and a variety of statistical tools. For training and validation, they use the current and past history of input data and feedback outputs via unit delay lines.

Network architectures that are used for RUL estimation can be classified as being either (i) feedforward (static) networks in which the inputs to a particular layer depend upon the output of the preceding layer. Or (ii) dynamic networks, which the inputs to a particular layer are dependent on the outputs of the proceeding nodes as well as preceding iteration of the network itself. Static networks provide only one set of outputs from the model and their response to a set of inputs is independent of the previous model state (i.e. its last set of outputs). Most prognostic RUL models developed to date have been based on these static networks, with the multilayer perceptron being the most popular architecture (Sikorska, Hodkiewicz and Ma, 2011).

(Huang, Runqing, et al., 2007) used back propagation (BP) neural network methods for the prediction of a ball bearing's remaining useful life based on self-organizing map (SOM).

BP was also utilised with a grey model for predicting the machine health condition by (Dong et al., 2004). Time Delay Neural Networks are used for the prognosis of crack

propagation through aluminium by (Khan, Faisal, et al., 2015), with training data provided from a Virkler data set.

(Malhi, Yan, and Gao, 2011). Presented a competitive learning-based approach to long-term prognosis of machine health status. Specifically, vibration signals from a defect-seeded rolling bearing are pre-processed using Continuous Wavelet Transform (CWT). Statistical parameters computed from both the raw data and the pre-processed data were then utilized as candidate inputs to a Recurrent Neural Network (RNN). Based on the principle of competitive learning, input data were clustered for effective representation of similar stages of defect propagation of the bearing being monitored.

(Peel, 2008) won the IEEE GOLD category of the PHM psila08 Data Challenge. The task was to estimate the remaining useable life left of an unspecified complex system using a purely data driven approach. The method was based on the construction of Multi-Layer Perceptron and Radial Basis Function networks for regression. A Kalman filter provided a mechanism for fusing multiple neural network model predictions over time. A genetic algorithm was integrated to train ANN for predicting wind power under high uncertainty conditions (Ak et al., 2013). Feed-Forward Neural Networks (FFNN) were used to predict RUL for a bearing failure by (Mahamad et al., 2010).

3.8 Knowledge Based Models

The development of a mathematical model can be difficult to obtain in real-world situations due to limitations in the collected data or the complexity of the process. However, there may considerable statistical historical failure information available and from this knowledge-based models (KbM) or experience based prognostic approaches are often the simplest and cost effective method for the prediction of the RUL (Vachtsevanos et al., 2006).

Knowledge based models correlate the similarity between an observed situation and a databank of previously defined failures and deduce the life expectancy from previous events and replicates the way in which human expert solves problems. Sub-categories include the following: a. Expert systems. b. Fuzzy systems.

3.8.1 Expert systems

An expert system is a software program (or set of programs) consisting of a knowledge base containing accumulated experience from subject matter experts and a rule base for applying that knowledge to particular problems known to the software system. Rules are formulated as precise IF-THEN statements; these are often based on heuristic facts acquired by one or more experts over a number of years (Sikorska et al., 2011). To be useful, a knowledge base must be as complete and exact as possible (Biagetti and Sciubba, 2004).

Expert system suffer from certain disadvantages; firstly the system itself is only as good as the experts developing the rule base, secondly there is no ability for the system to provide confidence limits, thirdly as the rule base increases, so does the computational complexity (e.g. combinational explosion problem) and lastly, the system's ability to handle new situations that are not defined in the rule base. Furthermore, being determined by a discrete set of rules, production of a continuous variable prediction is unfeasible and limits the usefulness of RUL estimation directly. However, using the rules, an expert may be able to infer discrete RUL estimations in terms of days, week etc.

An expert system was presented by (Biagetti and Sciubba, 2004) that incorporated a knowledge base, which was able to predict both faults and perform both diagnostics and prognostics for energy conversion plant. Although unable to predict an RUL, IF-THEN rules were defined with estimated fault indicators, and Fuzzy representation was used for fault diagnosis. Prognostics was performed by projecting the future state based on the present state and foreseen operative conditions.

3.8.2 Fuzzy systems

Expert systems use classic predicate logic, a statement can be either true or false; data can thus be included or excluded from a set. It is not always easy however to define sets and associated membership so precisely. Fuzzy set theory overcomes this deficiency by allowing partial set membership based on a variable's 'degree of truth'. As in expert systems, fuzzy logic systems consists of a knowledge-base, fuzzy rules and implementation algorithms. These rules are descriptive but unlike expert systems, they are intentionally imprecise. A typical fuzzy process logic statement may look like

'IF (process is too hot) AND (process is heating rapidly) THEN (cool the process quickly)' (Sikorska et al., 2011). Other methodologies such as neural networks (NN) or expert systems are usually used in conjunction with Fuzzy logic systems within literature.

(Choi et al., 1995) used an on-line fuzzy expert system, called alarm filtering and diagnostic system (AFDS), to provide the operator with clean alarm pictures and system-wide failure information during abnormal states through alarm filtering and diagnosis. In addition, it carries out alarm prognosis to warn the operator of process abnormalities.

(Amin et al., 2005) used an in-line health monitoring system developed for hydraulic pumps that utilize feature extraction, fuzzy inference systems, and knowledge fusion. These techniques were applied to hydraulic pump diagnostics in order to create a robust system for the determination and classification of pump degradation.
(Chinnam et al., 2004) propose in the absence of degradation and/or adequate failure data, domain experts with strong experiential knowledge, can be used to establish fuzzy inference models for failure definition. The paper presents a neuro-fuzzy approach for performing prognostics under such circumstances. The proposed approach is evaluated on a cutting tool monitoring problem.

The generalised advantages and disadvantages of the main prognostics techniques e.g. PoF, data driven and Knowledge based approaches investigated from the literature review above are presented in table 3.1. (Sikorska et al., 2011) presents a much more comprehensive overview of each technique breaking down within these main headings.

	Advantages	**Disadvantages**
Model Based (Physics of failure)	Estimates are the most accurate and precise of all modelling techniques. Provides confidence boundaries. Easy to understand outputs. Dynamics of the states can be estimated and predicted at each time	Expensive to develop Complete and high level of knowledge of the component behaviour needed Often uses experimental data to develop a model, which depends upon conditions such as environmental or material Can be component or system specific

Data Driven	Lower cost Greater tolerance to changes in environmental, material, usage Can be implemented at all system levels Easier to construct and generally simpler Flexibility in adaptation. Well established techniques	Highly reliant upon the data, particularly run to failure data, which is not always available. Failure thresholds may be hard to establish. High computational loading may be required (e.g. NN). Large data sets may be needed for training. Less precise
Knowledge Based	Simple, but can be time consuming to develop. Easy to understand Fewer rules required than for expert systems (Fuzzy) Inputs may be imprecise, noisy or incomplete (Fuzzy) Confidence limits can be provided on the output with some types of models (Fuzzy).	Domain experts required to develop rules Large number of rules required. Discrete outputs only (Expert) Significant management overhead to keep Knowledge base up to date. Precise inputs required. No confidence limits supplied. Not feasible to provide exact RUL output.

Table 3.1. Outlining advantages and disadvantages of the main prognostic techniques (Sikorska, Hodkiewicz and Ma, 2011).

3.8.3 Hybrid Prognostic Models

Hybrid prognostic approaches attempt to utilise the advantages from a combination of the three main prognostic approaches discussed above to enhance prognostic results. In all reality, it is quite rare that an individual prognostics approach will be entirely data-driven, model-based or knowledge based. For example, a model-based approach will incorporate some aspects of data-driven approaches; e.g. in the tuning of the model using actual condition data and data-driven approaches will often make use of model derived data.

As well as hybrid; fusion, combination and integration are all term banded in literature to represent this approach. Hybrid techniques, although starting to appear in literature with more prevalence, are relatively new, but show increasingly promising results for prognostics. The permutations are large and the main combinations with relevant examples are looked at below.

(Liao and Kottig, 2014) have produced a comprehensive review on some of the state-of-the-art techniques appearing in literature. The review below attempts to build upon this, especially looking for suitability for the applicability for relay health monitoring.

According to the findings, the hybrid approaches may be categorised into the following five types:

- H1 –Knowledge-based model + data-driven model
- H2 – Knowledge-based model + physics-based model
- H3 – Data-driven model + data driven model
- H4 – Data-driven model + physics-based model
- H5 – Knowledge-based model + data-driven model + physics-based model

The way the three categories combine is summarised in figure 3.10 below.

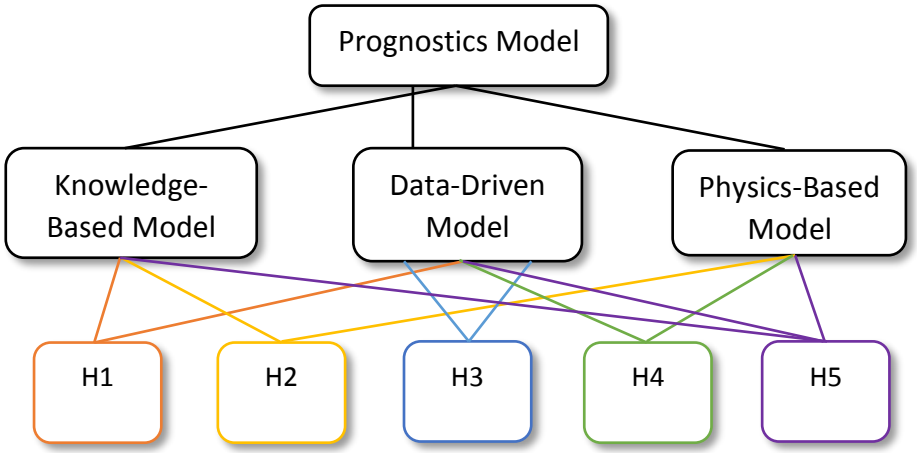


Figure 3.10. Hybrid model combinations (Liao and Kottig, 2014).

3.8.4 Knowledge-based model + data-driven model

From table 3.1 above, experience-based models alone using domain knowledge have the disadvantage of generally outputting discrete events and therefore cannot deal with continuous variables. This means that RUL predictions are very limited or restricted to discrete events. Fuzzy logic has the enhanced ability to deal with continuous variables via fuzzification using appropriately designed membership functions. Outputs can be converted into a numerical result, again using a corresponding set of membership functions. However, an expert is still relied upon to develop the underlying system on which the fuzzy representation is based upon.

Data-driven models however, can handle continuous data and have the ability to learn the structure and pattern from only the data. Therefore, H1 models have the advantages of integrating expert knowledge into data-driven models, thus allowing for system state estimation and prediction of RUL.

The data-driven component is usually achieved by the use of artificial neural networks ANNs which are combined either with the expert system of fuzzy logic system. A neurofuzzy (NN & FL) combination was applied to a gas turbine engine by (Brotherton et al., 2000). (Zhao et al., 2009) verified the effectiveness of the neuro-fuzzy model in predicting the health condition of bearings. Simulation and an experiment were carried out to verify the model, with results showing that the neuro-fuzzy model is a reliable and robust forecasting tool, and more accurate than a radial basis function network alone. (Wang et al., 2004) compared Recurrent Neural Networks (RNNs) and Neuro-Fuzzy (NF) systems. The performance of these two types of predictors is evaluated using two benchmark data sets. Through comparison it is found that if an NF system is properly trained, it performs better than RNNs in both forecasting accuracy and training efficiency. Accordingly, a NF system was adopted to develop an on-line machine fault prognostic system. (Gaga, 2001) used a hybrid reasoning approach that is capable of integrating domain knowledge and test and operational data from the machine. This approach was illustrated with an industrial gearbox example. In this approach explicit domain knowledge was expressed as a rule-base and used to train a feedforward NN. (Satish and Sarma, 2005) combined Neural Networks and Fuzzy Logic and formed a Fuzzy Back Propagation (Fuzzy BP) network for identifying the present condition of the bearing and estimate the remaining useful time of the motor.

3.8.5 Knowledge-based model + physics-based model

The integration of knowledge-based models and physics-based models is often based on the knowledge-based model being used to enhance the performance of the physics-based model.

Signal processing and neural network tracking techniques, along with automated reasoning, classification, knowledge fusion, and probabilistic failure mode progression algorithms was used by (Byington, 2004) to predict the health of aircraft actuator components. The level of damage (damage index) was quantised by a fuzzy logic implementation utilising a predefined set of rules tailored from knowledge of the system and engineering judgment. Kalman filtering was then applied to predict the progression of the damage.

Crack growth in tension steels bands was looked at by (Swanson, 2001), who developed a prognostics methodology with a Kalman filter and fuzzy logic. Using vibration mode frequencies as responsive fault features, the Kalman filter performed

feature tracking and forecasting, and the remaining useful life was estimated using Newton's method. Fuzzy logic was used to adapt failure thresholds to the operational conditions.

3.8.6 Multiple data driven models

This category consists mainly of two main approaches. Firstly, the systems degradation state is estimated by one data-driven model and the future RUL by is then determined by another data-driven model. Or secondly, multiple, competing data-driven models may be used to predict RUL and their results are then aggregated to improve the prediction performance by fusion.

(Liu, 2012) presented an approach which combined least squares support vector regression (LSSVR) with hidden Markov models (HMMs) for bearing remaining useful life prediction.

(Huang et al., 2007) predicted the remaining useful life of bearings combining two data-driven approaches. The systems degradation was calculated by the minimum quantization error (MQE) of the established baseline self-organizing maps trained with vibration features which established the health index for the system. The health index was then used as input to back propagation neural networks for remaining useful life prediction.

The tool wear condition in drilling operations was estimated by (Yan and Lee, 2007) via logistic regression analysis combined with a maximum likelihood technique (for parameter estimation), based on features extracted from vibration signals using wavelet packet decomposition. Another data-driven model, which was an autoregressive moving average (ARMA) model, was applied to predict the remaining useful life based on the estimated tool wear condition.

3.8.7 Data-driven model and physics-based model

Approaches combining data-driven and physics-based models have been studied extensively in literature with the enhancement of prediction performance.

The use of a data-driven model to infer a measurement model, with a physics of failure model to predict remaining life has been used extensively to provide remaining useful life estimates for electronics systems by CALCE (Kumar, Torres, Chan, and Pecht, 2008), (Zhang, Kang, and Pecht, 2009), (Patil, Das, Yin, Lu, Bailey, and Pecht, 2009), (Cheng and Pecht, 2009), (Pecht and Jaai, 2010) and (Xing, Williard, Tsui, and Pecht,

2011). In these papers the data-driven model is used to estimate the system state and by using a healthy state reference, system anomalies are detected by comparison techniques such as Mahalanobis distance. Physics-based methods are then used to perform RUL prediction by projection using techniques such as Kalman and Particle filtering.

(Mohanty et al., 2008). Used a hybrid prognostics model which combines the information from off-line data-driven and physics-based, and on-line system identification-based predictive models. The model was built for real-time remaining useful life estimation of metallic aircraft structural components.

Many other papers have been produced by using a data-driven model to infer the measurement model, the measurement model is used for mapping the sensory data to underlying system state, which is not measured and a physical model is used to predict RUL. (Orchard and Vachtsevanos, 2007), (Rosunally et al., 2009), (Peng et al., 2012), (Saha et al., 2007), (Baraldi et al., 2012.).

One Data-driven model and physics-based model that has particular relevance to this body of work is on structural damage to BGA interconnects incurred during vibration testing has been monitored in the pre-failure space using resistance spectroscopy based on state space vectors, rate of change of the state variable, and acceleration of the state variable by (Lall, Lowe and Goebel, 2010). The future state of the system has been estimated based on a second order Kalman Filter model and a Bayesian Framework. The technique was extended by using a non-linear model of degradation and the future state of the system has been estimated based using a second order Extended Kalman Filter model and a Bayesian Framework (Lall, Lowe, and Goebel, 2012). Extended and Unscented Kalman filters allow the extension of the traditional Kalman filter to be able to deal with systems that are not bounded by the constraints of linearity. In the case of the Extended Kalman filter by linearizing about an estimate of the current mean and covariance and for the Unscented Kalman filter by approximating the probability distribution by a deterministic sampling of points which represent the underlying distribution as Gaussian. A nonlinear transformation of these points is used as an estimation of the posterior distribution, from which the moments can be derived.

A data-driven model may be used to replace a system model in a physics-based model. The lack of an analytical degradation model can be compensated by replacing this model with a data-driven technique, where there is not an explicit analytical

equation for modelling of the physics. The parameter estimation process is similar to the framework of the filtering method (e.g., Kalman filter, particle filter), (Liao and Kottig, 2014).

In the paper by (Chen et al., 2011) the physic-based model usually used to model the degradation in the prognostics approach was replaced by an Adaptive Neuro Fuzzy Inference System (ANFIS) which was trained with historical data to model the degradation process. The technique was used for the prognostics of crack growth in helicopter gear plate and bearings. The trained ANFIS was used to construct an n^{th} order hidden Markov model within a high-order particle filter approach. The particle filter carried out n-step ahead predictions to extrapolate the system fault indicator and remaining useful life probability density functions.

A third method in this category is the use of data-driven models to predict future measurements, and use a physics-based model (PbM) to predict RUL. PbMs are often used conjunction with Baysian update methods to update the current model with new measurements. When generating n-step ahead predictions for RUL estimation, using particle filter as an example, acquisition of new measurement is not available. This can lead to low prediction accuracy in many applications because the identified model during the state estimation period may not be accurate and robust.

(Liu et al., 2012) used a Physics based Particle filter for system state prediction together with an integrated data-driven prognostic. The particle filter handled system state estimation and parameter identification based upon Bayesian learning. A data driven predictor (ANN based) was used to predict the future measurements and this was incorporated into the particle filters learning structure. This prediction formed the new measurement to update the weights of the particle filter, thus improving the long term remaining useful life estimation. This approach was used in comparison with a single prognostic approaches of a particle filter alone and data driven predictor to demonstrate its enhanced accuracy in RUL prediction of Lithium-ion battery discharge.

Lastly within data-driven and physics-based models is the fusion of their results. This type of hybrid approach simultaneously predicts a system's RUL with two prognostics models, i.e., a data-driven model and a physics-based model. The final RUL prediction result is calculated by fusing the results of the two prognostics models. It can improve the prediction accuracy, and narrow the confidence boundaries, (Liao and Kottig, 2014).

(Goebel et al., 2006) used a fusion methodology for the prognostics of bearings. A physics based model of the underlying system in this case bearing spall growth was fused with pairs of data measurements from speed and load data and health states for learning a spall length model. The predictions of both estimators and their results are then combined in order to improve the RUL prediction. This work was continued to employ Dempster-Shafer regression to negotiate the different damage estimate (Goebel et al., 2007.)

3.8.8 Experience-Based Models, Data-Driven Models and Physics-Based Models

Prognostics strategies in this category may be extremely difficult and impractical to implement due to the difficulty that might be encountered by each type of model (experience-based, physics-based, or data-driven models). The benefits however are potentially huge, with the ability to take all the strengths of the previously discussed techniques and fuse all types of information (e.g., domain knowledge, maintenance feedback, and condition data and physics), (Liao and Kottig, 2014). A lot of problems still revolve around how to incorporate results and fuse together each type of model.
A prognostic fusion model built upon an optimal linear combination of the RUL estimates of single prognostic algorithms by using the absolute value and prediction error as the index of prognostic precision was presented by (Xu and Xu, 2011).
(Gola and Nystad, 2011) presented a diagnostic and prognostic scheme for assessing the health state of a choke valve under erosion, and estimating its remaining useful life and (Orsagh et al., 2003) used model-based information to predict the reaming useful life of bearings with bearing spall initiation and propagation models.

3.9 Condition Based Monitoring for Electrical Contacts

Analysis of electrical contact wear has been the subject of research for many years; however over the last eight years, research has started to appear on the condition monitoring within relay contacts. A brief review of prior work is summarised below.

In the work carried by (Yao, Fang, et al., 2006) on the dynamic contact resistance, measurement devices and test methods are firstly introduced. Secondly, the main factor of analysis is reviewed and three diagnostic parameters are extracted. Thirdly, both statistics analysis and sequence analysis are looked at. The analysis of the

results indicates that the statistical characteristics and sequence tendency can predict the electrical contact reliability. A Fuzzy Overall Evaluation model of the electrical contact reliability is developed in the light of the fuzzy theory. By the model produced, the electrical contact reliability can be evaluated and predicted.

(Yu, Qiong, et al. 2009). Relates how the electrical life is an important index to evaluate the reliability of relay, and it is closely related with many characteristic parameters of the relay such as contact resistance, pick-up time, over-travel time, etc. By using a time series analysis and by taking some characteristic parameters as predicted variables, the life of relay can be obtained by the life prediction which is a non-destructive and short-cycling life test method. However, the paper shows the predicted accuracy is greatly influenced by the complex variations of characteristic parameters, and as a result it sometimes becomes too low to be accepted. For the purpose of improving the predicted accuracy, a life prediction method for space relays based on wavelet transform and ARMA (auto-regression moving average) time series is proposed. Taking the pick-up time of a relay (which can be regarded as a non-stationary time series) as an example, and by dividing it into three parts; trend term, cyclic term and stochastic term by adopting wavelet decomposition and reconstruction, the authors show models of each term can be built with different methods and then they are employed to predict the future values, respectively. Finally, the predicted results of the original time series are obtained by superimposing the respective prediction of each term.

The comparison between the method used in this paper and the conventional regression analysis method indicates that the former has higher predicted accuracy.

(Xuerong et al., 2010) states the traditional reliability assessment methods for electromagnetic relay are based on censored failure time data and provides very little reliability information. Furthermore, this method may be unavailable in applications with few or no failures. Actually, many time parameters of electromagnetic relay, such as closing time, over-travel time provide an obvious degradation process during the life test.

By testing time parameters of electromagnetic relay during its life circle, this paper presents a novel reliability assessment method of electromagnetic relays based on multiple time parameters degradation and multiple failure mechanism. Considering the difference of electromagnetic relay degradation, regression analysis method and time

series method are employed to build the degradation model of time parameters for the electromagnetic relay.

Multiple degradation parameters form a reliability assessment model which can evaluate the reliability of all failure mechanisms, which is developed with stochastic theory. By analysing the failure mechanisms of the electromagnetic relay, this paper presents a novel reliability assessment method of electromagnetic relay based on multi-time parameters degradation and multi-failure mechanism. The degradation process of electromagnetic relay is described with a two dimensional stochastic method, and the electromagnetic relay reliability assessment under material loss failure mode, gap bridging failure mode and insulation contamination failure mode is achieved. Compared with present reliability estimate method, the method proposed in this paper takes parameters interaction and detail degradation into account. So, it is a more reasonable method for electromagnetic relay reliability assessment, as well as the reliability assessment of some other components.

(Xuerong, Ye, et al., 2012). Considers the effect of the environment on the reliability prediction, stating post-analysis methods do not consider the influence of different application environments on electromagnetic relays, only using coefficients to revise the prediction results. In fact, the failure process and failure mechanisms of electromagnetic relay may totally change along with the environment. The paper collects the failure process data of electromagnetic relays, and studies the degradation characteristics under different failure modes, finally establishing and proposing the regression degradation model and the failure physical degradation model of relay contacts respectively, and verifies the accuracy through the life prediction based on the model.

The conclusion drawn from this paper, show further studies on the establishment of a more accurate regression degradation model that takes into account physical degradation. A physical degradation model of the abrasion failure, as well as the bridging failure and the contamination failure, and a method of the parameter estimation are then used to predict and value the life of the electromagnetic relay samples in four steps.

Firstly, analysis of the sensitivity of the different degradation parameters to the contact performance under different failure mechanisms, the over travel time and the rebound

duration, the over travel time, the fluctuation coefficient respectively as the predicted variables of the abrasion failure, the bridging failure and the contamination failure.

Secondly, the proposal of the pre-processing method of how to deal with the predicted variables based on the wavelet transform and the stationary time series analysis.

Thirdly, it takes advantage of the linear regression analysis method, to establish the linear degradation model which regards the operation time as the independent variable and the predicted variables of the failure mechanisms as the dependent variable. This is used to derive the equation for the life prediction, the accuracy of the predicted results, and the reliability evaluation.

Lastly, from the influence factors of the material transfer to the contact gap, the combined influence of the arc energy the contact surface morphology to the degradation rate of the contact gap, and use of fatigue cumulative damage theory is used to establish the failure physical degradation model of the electromagnetic relay contacts. This model can efficiently characterize the influence of the environment condition and the product structure to the degradation process of the electromagnetic relay contact performance.

The work carried out in (Yao, Fang, et al., 2012) proposes the analyses of the uncertainty of bounce time of contacts for relay and its use in predicting operating reliability. It changes the bounce time series into symbolic series according to the threshold function given in the paper. By means of the coding symbolic series and probability statistical analysis, the use of a concept named series entropy, which is a suitable Eigen-value for temporal schema and helps to find the time-change law of uncertainty of bounce time series. The analysis indicates that series entropy of bounce time for bad contacts descend as time goes on. The law can be used to predict the operating reliability.

The sequence of symbols is applied to describe the uncertainty of bounce time of the electromagnetic relay in the closing process. According to the entropy *Hs* which is obtained from the symbol sequence, the state of the electromagnetic relay can be assessed. For the chaotic or stochastic data, sequence entropy will tend to be 1 or has waves with ups and downs around a high level; for the deterministic data, sequence entropy will tend to be zero or has a persistent downtrend till to failure;

Although the value of the sequence entropy which ranges from 0 to 1 is indeterminate and uncertain, there is a definite tendency just before its failure.

3.9.1 Prognostic and Reliability for Electrical Contacts

A prognostic model for separable electrical contacts is considered by (Ostendorf, 2014). The paper examines providing reliability to separable electrical contacts and looks at the mandatory measures for the discrete nature of the contact interface as well as the necessity of fulfilling a broad variety of product requirements. The paper also examines the formation of real and conductive contact area controls for the reliability and efficiency of an electrical contact. These processes depend on a great number of independent or interrelated factors. The variety of these factors can be divided into (a) performance factors governed by the operating conditions and (b) design-technological factors determined by fabrication characteristics of the contact unit. The paper derives a model to determine the influences of the design technological factors based on the reliability and quality of electrical contacts. The consideration of mathematical models for reliability and failure such as Weibull, Arrehenius, Hallburg-Peck and Coffin-Manson. Together with the lifetime tests and their statistical evaluation it is possible to deduce fault rates combined with the probability of failure and reliability for a newly designed contact unit. Furthermore detailed surface analytical investigations of the contact zone were performed to identify the occurrence of physical and chemical processes which are influencing the state of the contact interface and finally, the contact resistance and reliability.

A considerable body of work has been directed around the failure of electrical connectors used in integrated circuits and associated static connections subject to temperature and environmental stresses such vibration.

(Lopez et al., 2008) presents a methodology based on the physics of failure and the sequential probability ratio test, for modelling and monitoring electrical interconnects in health monitoring, and electronic prognostic applications. The resistance behaviour of an electrical contact was characterized as a function of temperature. The physics of failure of the contact technology was analysed. A contact resistance model was selected based on the a-spot radius and a function of temperature. The parameters are estimated as a power law and were fitted using the temperature characterization data using linear regression.

The physics of failure model was evaluated with a reliability application (temperature cycle test), and was found to produce estimation errors of < 1 m Ohm during a training period. The temperature and resistance of ten sample contacts were continuously monitored during the temperature cycle test, identifying the maximum temperature and resistances for each cycle. Using the physics of failure model, maximum resistance estimates were generated for each test sample. The residual between the monitored and estimated resistance values was evaluated with the sequential probability ratio test. The method was shown to overcome the issues of traditional threshold-based monitoring approaches, providing accurate resistance estimates, and allowing the detection of abnormal resistance behaviour with low false alarm and missed alarm probabilities.

In (Lall, et al. 2012), again the leading indicators of failure has been developed to monitor the progression of fretting corrosion in electrical connectors and prognosticate remaining useful life. Connectors subjected to harsh environments may experience vibration resulting in fretting corrosion and degradation in contact resistance over time. Tin coated, rectangular-pin and socket electrical connectors have been studied. In this paper, a random vibration test profile has been used to stimulate the contact resistance degradation due to connector fretting corrosion. The contact resistance has been measured in situ using the resistance spectroscopy method in conjunction with phase sensitive detection. It has been shown that precise resistance spectroscopy and phase measurements can provide a leading indicator of failure significantly prior to the traditional definition of failure.

The prognostic approach here accepts the inputs to the system are not measureable and using the resistance measurement creates a feature vector model based on the resistance spectroscopy. The resistance change of 0.3 Ω has been used as a leading indicator of failure for the electrical connector prior to the traditional deformation of failure caused by fretting degradation. A Kalman filter was used as a recursive algorithm to estimate the true state of the electrical connector.

A technique has been developed for monitoring the structural damage accrued in ball grid array interconnects during operation in vibration environments (Lall, 2012). The technique uses resistance spectroscopy based state space vectors, rate of change of the state variable, and acceleration of the state variable in conjunction with Extended

Kalman filter, and is intended for the pre-failure time-history of the component. Condition monitoring using the presented technique can provide knowledge of impending failure in high reliability applications where the risks associated with loss-of-functionality are too high to bear.

The future state of the system has been estimated based on a second order extended Kalman filter model and a Bayesian Framework. The measured state variable has been related to the underlying interconnect damage using plastic strain. The performance of the prognostication health management algorithm during the vibration test has been quantified using performance evaluation metrics. Model predictions have been correlated with experimental data. The presented approach is applicable to functional systems where corner interconnects in area-array packages may be often redundant.

From the literature search it can be seen quite a large body of work has been developed on ascertaining the reliability of relays via various methods, however, little work has been carried out so far in producing a prognostic solution to switching contact failure. This research proposes to remedy this, by suggesting a model based prognostic solution of the most common failure mechanism within the relay, namely, the electrical contacts.

3.10 State of the Art

From the Literature Review it can be appreciated that there is a large and growing body of literature on CBM and Prognostics. Methodologies produced so far however, seem to be application specific, for example PbMs to determine crack extension, or DD methods from specific data sets. The successful applications of prognostics to complex engineering systems is still within its infancy; problems of how to model the high non-linearity and stochastic behaviour which contribute to degradation is still a major problem, however, this is steadily being resolved.

The application of prognostics to electrical contacts is still to a large extent driven by knowledge based model assumptions. Statistical and sequence based models are used in numerous reports to infer statements of lifetime and the reliability of the electrical contact. These statements must in turn be replaced or described with statements on their probability.

This approach is the simplest, the process of deriving complex models, is replaced by knowledge based assumptions founded upon historical data for RUL prediction and is commonly used by manufactures to give an indication of failure time. The most recent application of this to prognostics is the approach by (Ostendorf, Wielsch, and Reiniger, 2014) in which, together with lifetime tests and their statistical evaluation, it was found to be possible to deduce fault rates combined with the probability of failure and reliability for a newly designed contact unit. This approach was developed from knowledge based applications from FMEA on comparable products as well as experience from the development process. The compilation of all potential damage patterns for all of the constituent components leads to an assessment of which fault is likely to occur based on experience. From this assessment, distributions can be derived and along with stress failures that are also taken into account, their influences are quantified in terms of an increased theoretical error resulting in the previously defined damage causes.
Mathematical models are used as descriptors for the reliability, namely Weilbull. The stresses being modelled by Arrhenius, Coffin-Manson and Hallberg-Peck, with accelerated testing being used to derive the parameters for these models. The final prognosis of error and failure rates were ascertained with the aid of Weibull parameters, which were determined beforehand in experiments, in consideration of the acceleration factors.
Where this approach offers a simplified methodology, it relies heavily upon expert system knowledge, and with the models being defined from past data, potential new situations could pose problems, for example changes in the operating environment.

Data driven methodologies in summary utilise regularly collected condition monitoring data to produce a system model instead of physics based or expert knowledge. Both statistical and machine learning (A.I) techniques are headed under this category. A.I based techniques attempt to recognise complex patterns in the data and utilise them to make intelligible decisions; where in statistical based methodologies, models are constructed by fitting a probabilistic model to the available data. One noted problem from the literature review is that to infer an accurate degradation pattern model from an A.I based approach can utilise a large amount of failure data. This data may be difficult to obtain in large quantities, as in general, systems are not run to failure and in some systems this may take months/years to evolve. This leads on to the problem

of determining failure thresholds, as within the data, it may not be apparent what the overall failure mechanism consists of.

The work done by (Yu, Qiong, et al. 2009), forms a basis for using time series analysis, utilising characteristic parameters as predicted variables in electrical contacts. The life prediction method for space relays based on wavelet transform and ARMA (auto-regression moving average) time series proposed, offers a statistical DD methodology. This method however still utilises past data and does not have the ability to adapt to online changes to condition monitoring data from unforeseen events. The use of the ARMA model is however an attractive solution due to its low computational overhead and comparisons between the ARMA method and conventional regression analysis (Xuerong et al., 2010). (Xuerong, Ye, et al., 2012) indicates that the former also has a higher predicted accuracy. It is noted that work above is more of a reliability prediction tool, as it offers no way of predicting impending operating environment, future load profile uncertainties and input data uncertainties such as the initial state of the system, variability in material properties and manufacturing variability. Also no metrics are provided in order to bench mark the techniques in terms of RUL prediction.

The use of physics based models in prognostics for electrical contacts is small and this is perhaps due in part to the complexity of the physical process relating to the system degradation. The need for failure based degradation data is less important in physics based models, however an understanding of the physical process that accounts for the degradation is essential and can be very difficult, if not almost impossible to determine. Physic based modelling constitutes two phases, the first being to obtain the residuals that represent the deviation of measurements from the expected values for a healthy system and secondly the mathematical modelling of the failure degradation. The challenges that arise from physics based approaches incorporate the need for a knowledge of the physics that cause the degradation, as well as a thorough understanding of the component/systems and how environmental and operational conditions could play a significant part/dimension.

One such use of a simplified physics model was made by (Lall, et al. 2012) in monitoring the progression of fretting corrosion in electrical connectors. State space vectors, incorporating rate of change of the state variable, and acceleration of the state variable was used as a recursive algorithm to estimate the true state of the

electrical connector. The future state of the system has been estimated based on a second order extended Kalman filter model and a Bayesian Framework.
In terms of a PbM, work by (Lopez et al., 2008) presents a methodology based on the physics of failure. A contact resistance model was selected based on the a-spot radius and a function of temperature.

3.11 Research Gap

This chapter provides the following conclusions:
The State of the Art methodologies in prognostics for electrical contacts are still very much based around knowledge based techniques such as mathematical model descriptors for the reliability, namely Weibull or regression based.
Due to the nature of the problem outlined in Chapter 2, relay failure modes are extremely complex, consisting of multiple parameters which superpose together. Ideally, a physics based model would be the perfect solution, being able to account for all eventualities such as**,** operating environment, future load profile uncertainties and input data uncertainties such as the initial state of the system, variability in material properties, manufacturing variability, etc.
On the other hand, one of the problems exhibited with data driven methods is the need for training data in the case of ANNs and statistical data to infer behaviour from. This may not always be available, and to develop a realistic model for a complex system, a great deal of different data would be needed.
Hence, a gap appears for a methodology that can develop a model from the data in real time and be able to produce prognostic prediction of RUL. This would enable the operating environment, future load profile and input data uncertainties to be catered for based on recent data history and the formation of a model that corresponds accordingly.

3.12 Contribution of this Thesis to knowledge

This PhD research contributes to literature by forming an accurate experimental data set for an electromagnetic relay contact extracted from accurate measurements of an attributed failure metric, namely contact resistance. This

benchmark data set is sufficient in size to allow a data driven approach to be applied.

The development of a real time data driven prognostic approach based upon modelling of degradation using system identification techniques is exploited. A sliding window recursive least squares approach is proposed that is able to adapt to operating environments, future load profile uncertainties and input data uncertainties, such as the initial state of the system, variability in material properties, manufacturing variability and component replacement. This prognostic has been benchmarked to enable an assessment of effectiveness to be ascertained.

In addition, a simple physics based model (PbM) based upon arc erosion of the contact and its effect on the contact resistance due to the reduction of contact surface area is derived.

The PbM of the arcing and the erosion process has been developed to enable mass loss to be calculated, the results are used in a state space model. This model enables the primary effects from arcing damage to be examined. The model incorporates the ability to look at heat flow through the contact, voltage across the contact due to heating and mass loss from the contact. The development of a model for the estimation of secondary parameters within the relay such as coil current, contact bounce and contact position has also been established. This state model was then used with particle filtering to enable a prediction of the RUL to be estimated.

Chapter 4 - Contact Resistance Data Acquisition

Introduction

This chapter focusses on the methodology used for the data extraction of a suitable feature vector for condition based monitoring (CBM) as well as the development and proving of a prognostics solution. The consideration of a suitable metric is first explored which reflects accurately the degradation of the relay contacts. This is followed by a discussion of the experimental apparatus devised and constructed within the laboratory to collect the data, including the quantification of measurement and process noise. Lastly, the raw data is discussed in terms of patterns and features that may be applicable for prognostics as well as filtering methodologies to assist with trend clarification.

4.1 Experimental Derivation

From Chapter 2 it becomes apparent that contact failure is the primary failure mode and is in the main part due to arc erosion and material transfer between one contact and another. This is due to the discharge of the arc causing contact surface damage, degradation of the performance, as well as eventual total failure of the contact. Failure modes may be dependent upon numerous factors; including the arcing time, loading, contact material, bounce and atmosphere and thus affects the amount of material transfer, shape and direction.

Three main categories of contact failures can be identified; failure from material loss, insulation contamination failure and gap bridging (Balme, 1990), the failure mode is also related to changes in contact gap. Contact surfaces are, relatively speaking smooth to the naked eye when under normal conditions and the contact gap travel is consistent. Once material transfer starts to occur, there is a loss of material from one contact causing the surface to be eroded, with material being deposited onto the opposite contact over a wide area. This may now cause an increase in the contact resistance as well as the size of the contact gap, which inevitably leads to reduction in the over travel time. (Morin, Laurent, et al., 2000), (Morin and Laurent, 2000).

Eventually, a point is reached where the over travel time becomes zero and the pair of contacts hardly touch one another, resulting in failure due to non-closing caused by the material loss. This mode of failure is defined as material loss failure. The other

extent of this is where the material transfer is concentrated over a small area as opposed to the whole contact, resulting in the formation of a small pip in one contact (usually the anode). This leads to the gap between the contacts eventually becoming zero and causes a failure mode known as bridging, where the contacts remain electrically closed even when opening.

In the third case, the situation may occur where the pip and the corresponding crater due to the material removal are more or less axially symmetrical and therefore the contact is gaining material at the same extent that its counterpart is losing material. This results in the equal shifting of the contact surfaces and thus the over travel distance remains constant. (Leung et al., 1991)

Lastly, a failure mode called insulation contamination failure can also occur. Contact resistance as the surface insulation contamination due to arc erosion increases also increase and can cause non-closing failure. The complexity of the how the contacts fail presents the question of what is the best degradation parameter to use in order to provide a reliable feature vector for condition monitoring. Numerous authors have examined various strategies which will be described below.

(Hammerschmidt et al., 2004) looked at the interaction between material transfer and contact kinetics yielding actual switching failures. This paper suggested that detailed information about the development of failure processes caused by material transfer may be gained by measuring the contact force in relays during life-tests. However, to measure contact forces the relay has to be modified, e.g. the fixed contact has to be removed and mounted on a movable force sensor. No modification is required if merely the values of both opening time and closing time are measured.

Arc duration and subsequent erosion measurements for contacts made of pure silver, silver alloys (AgNi, AgCu), and silver metal-oxides (AgCdO, AgSnOa) were carried out by (Jemaa, 1996) for switches and relays connected to complex circuits (motor, resistance, and lamp) used in automobile field. The actual parameters of the experiments such as the contact material, the environment, voltage and current (50 V, 0.1-3 A DC) and mechanics (opening speed) were controlled. The accurate voltage values of the consecutive arc plateaus included in arc phases are determined by statistical measurements and voltage histogram drawings.

The contact voltage drop, closing time and opening time were monitored by (Li, Kui, et al., 2000). In the end of test, the reliability of relay was estimated according to test data and the failure mode analysed.

Other measurements looked at are over-travel and contact gap (Xuerong et al. 2010), which was deemed extremely difficult to measure at present during the life of the test. The degradation and failure mechanism could only be analysed by testing time parameters of contacts, such as over-travel time and rebound duration. According to above mentioned analysis, different failure mechanisms will lead to different trends of contact gap and over-travel. The paper integrates over-travel time and rebound duration together to describe the degradation of contacts performance.

The most universally used metric for contact degradation in literature is contact resistance. This is also adopted by the manufacturers who often quote a contact resistance figure in their literature for a healthy contact and one deemed to have reached a point where failure is impending. This is useful, as it gives a threshold that may be used when a prognostic solution is proposed.

(Zhai, 2006) states contact resistance is an important parameter that directly reflects relay performance and it is also a basic datum to evaluate relay reliability. Based on theories of electromagnetic field, kinetics, contact force and electric contact, a full simulation analysis scheme of relay dynamic characteristics was presented by using the method of coupling finite element analysis and kinetic computation.

(Rieder and Strof., 1991) stated modern reliability requirements for relays cannot be satisfied unless the contact resistance is measured after each operation of the life test. A test device was developed to execute these measurements, and a special method was applied to reduce the resulting amount of data effectively without losing information. Commercial relays were investigated at intermediate and low power levels. Characteristic contact resistance patterns during the life of a relay were recorded depending on both the contact material and the electrical stress. Typical statistical patterns characterizing homogeneous and heterogeneous materials, erosion of contact plating, contact contamination were examined. (Chen, 1993), examined the contact morphology, surface composition, and contact resistance of DC relay contacts for two levels of low-current (less than 1 A) arc has shown that a similar contact erosion mechanism and similar contact resistance degradation exist for 0.5 A

and 0.75 A switching a resistive load. Material transfer is attributed to ion sputtering during arcing, as is the degradation of contact resistance determined at different operating cycles, which appears to be influenced by both the contact morphology and surface contamination. A simple model is introduced and used to explain the process of arc erosion and contact resistance degradation during testing.

4.2 Measurement of Contact Resistance

From the survey of literature, contact resistance is the most universal metric in assessing how contacts degrade and although it is a secondary measurement, it takes into account the multiple failure modes discussed above, in 4.1. Although it is significantly smaller as compared with the overall circuit resistance, the changes in the contact resistance can cause significant malfunction of the device. This is because the contact resistance can vary significantly with the changes in the real contact area, contact pressure variations, resistive film non-uniformity, and other factors.

The work in this Thesis requires a solution to allow measurements to be taken in situ e.g. within the FADEC unit, and hence contact resistance offers a viable metric for on-board prognostics via non-invasive measurement (CAA directives).

The contact resistance to be measured is extremely small, a perusal of manufacturer's data sheets shows the contact resistance to be of the magnitude of a few milliohms. This means to accurately measure the changes, the sensitivity of the instrument needs to be in excess of measurement magnitude.

One way to do this is to use a Kelvin (four wire) resistance measurement, figure 4.1. In a normal resistance measurement, the effect of the leads that are in circuit loop becomes part of the measurement, so as well as the contact resistance, the lead measurement is added.

The four wire measurement uses an ammeter and voltmeter to measure the current through the contacts and the voltage across them, a simple application of Ohms law then gives the contact resistance.

$$R_{contact} = \frac{Voltmeter\ reading}{Ammeter\ reading} \tag{79}$$

Because the sense current to the voltage meter is so small, in the range of pA, the resistance of the measurement leads is negligible in the final result.

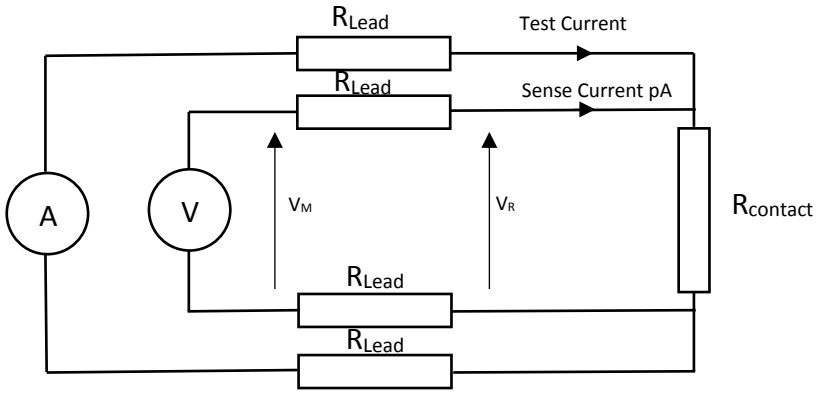


Figure 4.1. Showing the set-up of a four wire measurement.

The measurement of the contact resistance for this application needs to be taken when all the bounce has subsided, this is known as a static measurement as opposed to measuring the dynamic contact resistance which has been used be some authors, (Chen, Peng, et al., 2005).

4.2 Test Sample

Accelerated life testing is the process of testing a product by subjecting it to conditions (stress, strain, temperatures, voltage, vibration rate, pressure etc.) in excess of its normal service parameters in an effort to uncover faults and potential modes of failure in a shortened amount of time, (Wayne, 1980).

In order to accelerate the relay testing, various parameters were explored; in the end, it was decided to test the device at full voltage and current, ambient room temperature and maximum switching cycle specified for the device. This still resulted in test times in excess of two weeks in some cases. The test conditions that were used, tried to invoke as much as possible the actual in-use conditions for the relay. The notion to use inductive loading, that would have sped up the degradation process was initially revoked, as most relays are implemented with some degree of arc suppression. By using a resistive load, a benchmark may be established from which a comparison of contact life from inductive loading and increased arcing may be examined.

The relays specification is documented in table 4-1 below. From the table, the life of the relay contacts is in excess of 10^5 cycles with a resistive load and the mechanical components of the order of 2×10^6, initial contact resistance is below 100 mΩ (this takes into account the oxide that may be present on the contacts due to prolonged

storage or from manufacturing debris, hence the high initial value) and the maximum operating speed is 20 times/minute.

Characteristics	Item		Specifications
Contact	Arrangement		1 Form A
	Contact resistance (Initial)		Max. 100 mΩ (By voltage drop 6 V DC 1A)
	Contact material		AgSnO2 type
Rating	Nominal switching capacity (resistive load)		16 A 277 V AC
	Max. switching power (resistive load)		4,432 VA
	Max. switching voltage		277V AC
	Max. switching current		16 A
	Nominal operating power		400 mW (Standard type)
	Min. switching capacity		100 mA, 5 V DC
Electrical characteristics	Insulation resistance (Initial)		Min. 1,000 MΩ (at 500 V DC) Measurement at same location as "Breakdown voltage" section.
	Breakdown voltage (Initial)	Between open contacts	1,000 Vrms for 1 min. (Detection current: 10 mA)
		Between contact and coil	4,000 Vrms for 1 min.
	Temperature rise (coil)		Max. 55°C 131°F, Max. 45°C 113°F (200 mW type) (By resistive method, nominal coil voltage applied to the coil; contact carrying current: 16 A, at 20°C 68°F)
	Surge breakdown voltage (Between contact and coil) (Initial)		10,000 V
	Operate time (at nominal voltage) (at 20°C 68°F)		Max. 20 ms (excluding contact bounce time.)
	Release time (at nominal voltage) (at 20°C 68°F)		Max. 20 ms, Max. 25 ms (200 mW type) (excluding contact bounce time) (With diode)
Mechanical characteristics	Shock resistance	Functional	200 m/s^2
		Destructive	1,000 m/s^2
	Vibration resistance	Functional	10 to 55 Hz at double amplitude of 1.5 mm
		Destructive	10 to 55 Hz at double amplitude of 1.5 mm
Expected life	Mechanical (at 180 times/min.)		Min. 2×10^6
	Electrical (at 20 times/min.)		Min. 10^5 (at resistive load)
Conditions	Conditions for operation, transport and storage		Ambient temperature: –40°C to +85°C –40°F to +185°F; Humidity: 5 to 85% R.H.
	Max. operating speed		20 times/min. (at nominal switching capacity)

Table 4.1. Relay specification

The measurements of contact resistance were generated from an in-house developed test bed at Cranfield University. The test bed in figure 4.2 has been designed to allow continual cyclic testing of a relay at various currents and voltages, with the ability to vary the switching time. In-situ measurements of currents, voltages, temperature, pick-up time, over-travel time, the rebound duration, closing time and contact resistance are available for algorithm development. The test rig consists of the following major components: 1000 W resistive load, 50 A solid state relay, 16 A 30 V relay, bread board, data acquisition system connected to a computer. The prognostic rig is designed so that no other component will deteriorate quicker than the relay during the test. Each component is discussed below.

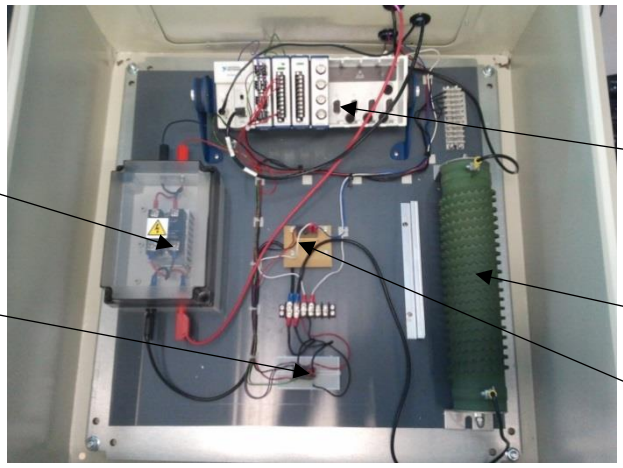


Figure 4.2. Relay test rig

Solid state relay: In order to allow a continuous current through the load in the test circuit and switch this current during testing, a 50 A solid state relay was applied due to its reliability under continuous operation. This was coupled with a heat sink to allow the dissipation of heat from the device.

Resistive load: In order to load the device in the test circuit, a suitable load needed to be employed. A ceramic bodied 1000 W, 1 Ω load was used to enable the current from the power supply to be directly proportional to that being switched be the relay under test.

Test Circuit Relay: a standard relay was used for switching the test circuit for the measurement. This relay allows the 6 V, 1 A test current to be switched in whilst the load current is switched out.

Bread board: The relay under test was mounted in rigid bread board as the connections are PCB type. As well as being PCB mountable, the relay also has TMP spade terminals. This allowed connection of the load circuit through the terminals and the voltage measurement to be taken as near to the contacts as possible on the bread board.

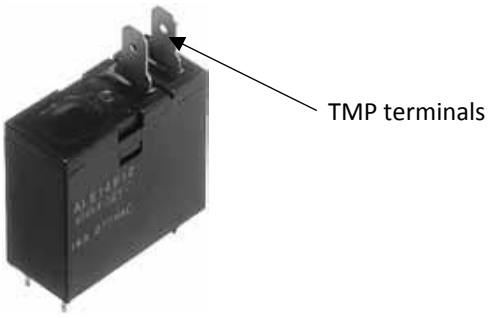


Figure 4.3. Picture of the test relay showing the TMP terminals.

Data Acquisition: National Instruments LabView® was used for the control of the test and data acquisition. The data collection was conducted with a NI 9219 Universal Analog Input, 24-Bit module and the outputs used for control of the test circuit, load circuit and relay under test was via a NI 9472 card which are connected to an NI cDAQ-9174 4-slot USB chassis.

For safety reasons due to the long periods the test rig would be left unattended and the potential burn hazard from the 1 Ω load, the whole unit was mounted on a PVC board and housed within a steel cabinet.

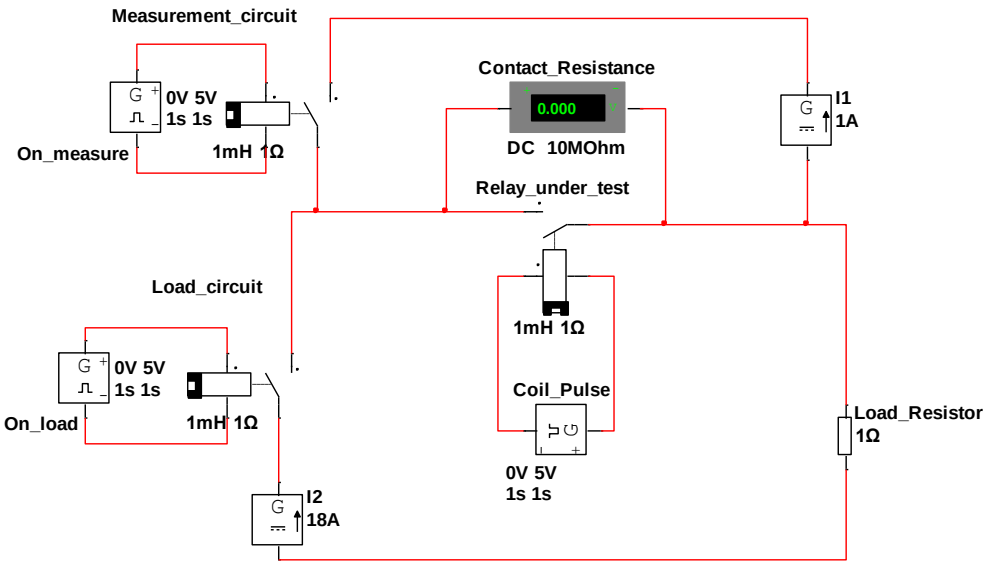


Figure 4.4. Test circuit for relay.

4.3 Test rig operation

The circuit diagram for the test rig is shown in figure 4.4 above. The main circuit consists of the load resistor connected to a power supply through the solid state relay and the relay under test. Across the relay under test, a second circuit containing a manufacturers specified test current of 1 A at 6 V is connected in series with a relay, this is the measurement circuit. The actual voltage reading is taken across the contacts of the relay under test.

The operation is as follows; the cycle time and number of cycles until test are entered into the LabView® display panel in figure 4.5. When the program is started, the load circuit is switched in via the solid state relay, the relay under test is operated at the required cycle time for the desired number of cycles, before a measurement is taken. Upon reaching this, the relay under test coil is kept energised, the load circuit is switched out and the measurement circuit is switched in. The relay under test, now has a 1 A current through the contacts and the voltage across the contacts is taken via the NI 9219. Because the current is 1 A, the voltage across the contacts is proportional to the contact resistance and this measurement is stored and plotted on the display.

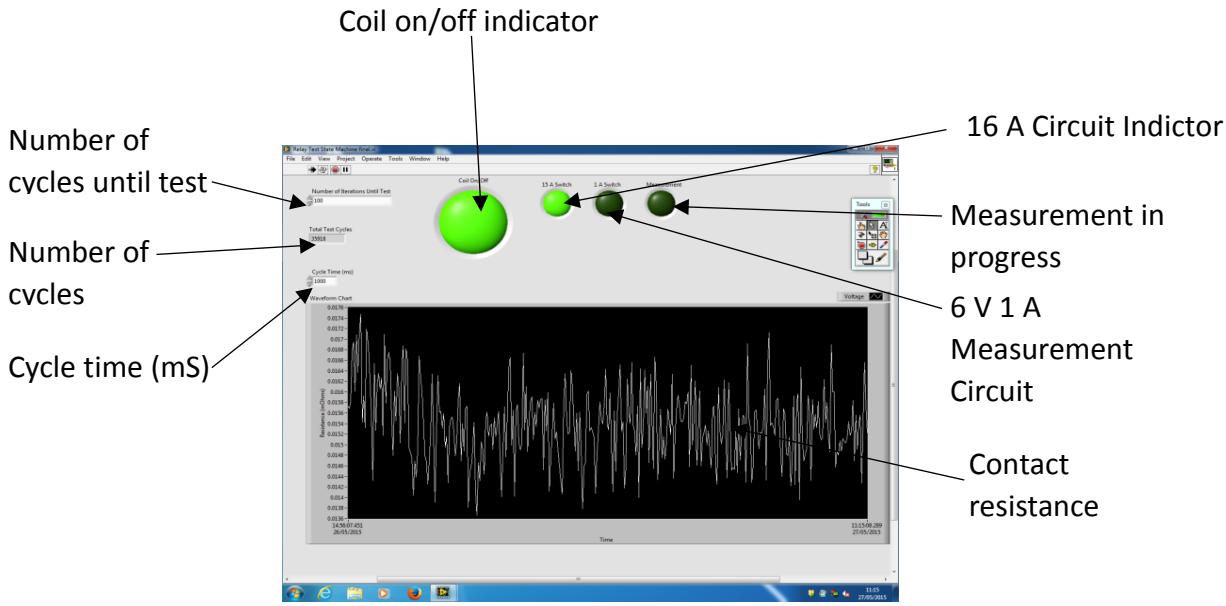


Figure 4.5. Labview display

A sample of seven relays were tested at their full load current and voltage within the same experimental conditions, and a reading of the static contact resistance every 100 cycles was taken until the End of Life (EoL). The EoL in each of the test samples,

resulted in a sudden failure due to the contacts no longer being able to separate (contact welding), and in each case occurred well past the manufactures stated operated point. The variation of the contact resistance between samples was interesting.

4.4 Measurement Uncertainty within the Results

In any set of experimental measurements, the reliability of the results will be influenced by uncertainties. The mathematical analysis of these uncertainties within a system and the variability of its parameters is given to an area of study called Uncertainty Analysis. The variability of these system parameters may be attributed to sources inherent to the system or from external disturbances and need to be quantified in any measured experimental results before solid conclusions can be drawn from the data.

A system is said to be linear if the corresponding values of its inputs (I) and outputs (O) lie on a straight line, such that

$$O_{ideal} = KI + a \tag{80}$$

where K is the slope and a is the intercept of straight line. In most cases the linear relationship is not obeyed causing non-linearity N(I). The system will also be influenced by variability arising from noise sources that are present in all systems, as well as noise added during measurement, for example the discretisation of the measured variables. These are referred to in measurement standards as Type A uncertainty errors that can be statistically evaluated from the set of measurement data (often considered as random uncertainty).

In general, the output (O) depends not only upon the input (I), but on external influences governed by environmental factors. Again in measurement standards, this is referred to as Type B error estimates influencing the measurements that are not directly observed from the measurement data (often considered as systematic uncertainty). These uncertainties on the system will be dependent on factors such as temperature, atmospheric pressure, power supply variation, vibration and humidity etc., which will all cause changes over time.

There are two main types of environmental effects or process noise as it often referred to. A modifying input causes the linear sensitivity of the element to change, the deviation I_M from I in a modifying environmental input causes a change in the linear

sensitivity from K to $K + K_M I_M$. Likewise, an interfering input $I_I$causes the intercept or zero bias of an element to change. If I_I is the deviation in an interfering environmental input from I_I in normal conditions, then this produces a change in zero bias from a to $a + K_I I_I$. The constants K_M, K_I are referred to in literature as environmental coupling constants or sensitivities.

Other parameters that are often considered within the systematic characteristics are hysteresis, resolution as well as wear and ageing. If the effects of hysteresis and resolution are deemed not to be present, then the input and output relationship can be assumed to be comprised of non-linear and environmental influences which may be represented by the general equation

$$O = KI + a + N(I) + K_M I_M I +, K_I I_I \quad (81)$$

Therefore, the aim of uncertainty analysis is to quantify, evaluate and combine all the various sources of uncertainty to produce a combined estimate of all the individual A and B type uncertainties within the system. This combined uncertainty $U_c$is estimated by a residual sum of squares (RSS) which gives a basic estimate, representing one standard deviation.

$$U_c = \sqrt{U_1^2 + U_2^2 + U_3^2 + \cdots + U_n^2} \quad (82)$$

The combination of uncertainties by the RSS technique assumes a standard relationship which is statistically independent. It is usually required for the above calculation to cover a larger population or condition, commonly 95% or 99% instead of the 68% pertained; therefore, scaling is employed to cover this larger population U_m.

$$U_m = kU_c \quad (83)$$

A type A uncertainty is determined by the statistical analysis of a series of observations (measurements) and will consist of uncertainties; including effects from variation of multiple repeated readings from the unit under test, effects of system noise and noise and short term variation of the standard used. In literature, this is deemed forward uncertainty propagation as opposed to inverse uncertainty quantification of the input (I) and model parameters based on a set of measurements of the output (O), an example of this being the comparison of a mathematical model with experimental output measurements.

4.4.1 Type A Measurement Quantification

In order to quantify the measurement error, a series of measurements were taken from a high accuracy voltage source, this being a Fluke 5700A multifunctional calibrator as recommended by National Instruments. This device is capable of delivering a low noise voltage output at within a range of 48 – 0.1 V with noise levels of ≤ 5 mV_{pk-pk} and ≤ 5 μV_{pk-pk} respectively. In carrying out the measurements, connections to the device are kept to the minimum length and tight, with a temperature of 23 ± 0.5 ºC and a relative humidity below 80 %. The NI 9219 is set up as in figure 4.6 below

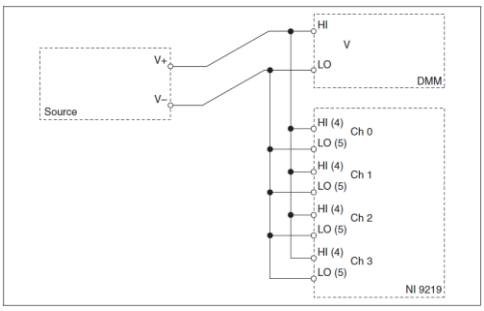


Figure 4.6. Calibration set-up

For each of the measurement values in table 4.2 below, 20 readings were taken and then averaged. This average is recorded as $V_{channel}$, and the deviation from the average for each measurement is then calculated.

Measurement Type	Max (v)
Voltage	3.2
	0.8
	0.1

Table 4.2. Calibration voltages

Figures (4.7), (4.8) and (4.9) show the distributions of the measurements for the 20 measurements taken at 3.2, 0.8 and 0.1 V respectively.

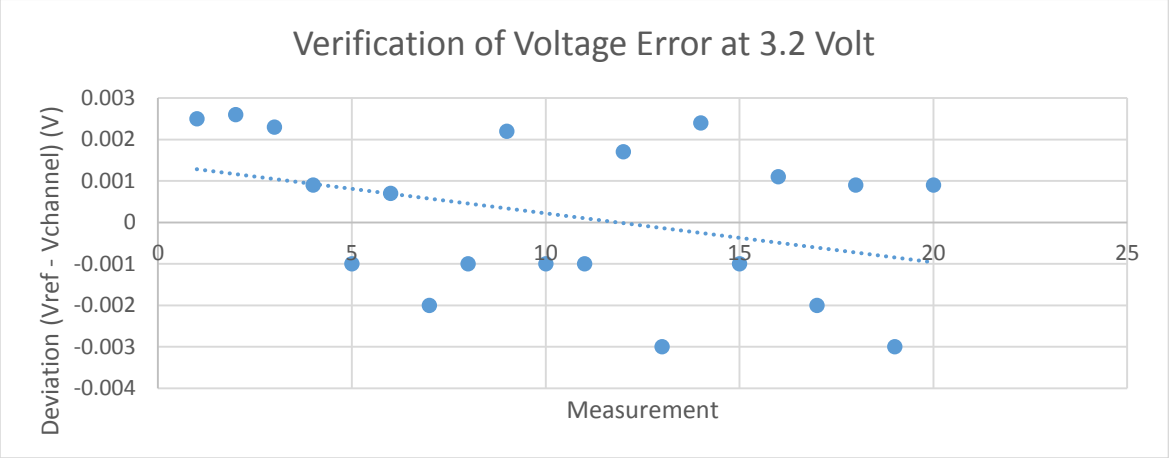


Figure 4.7. Voltage measurement distribution at 3.2 Volts

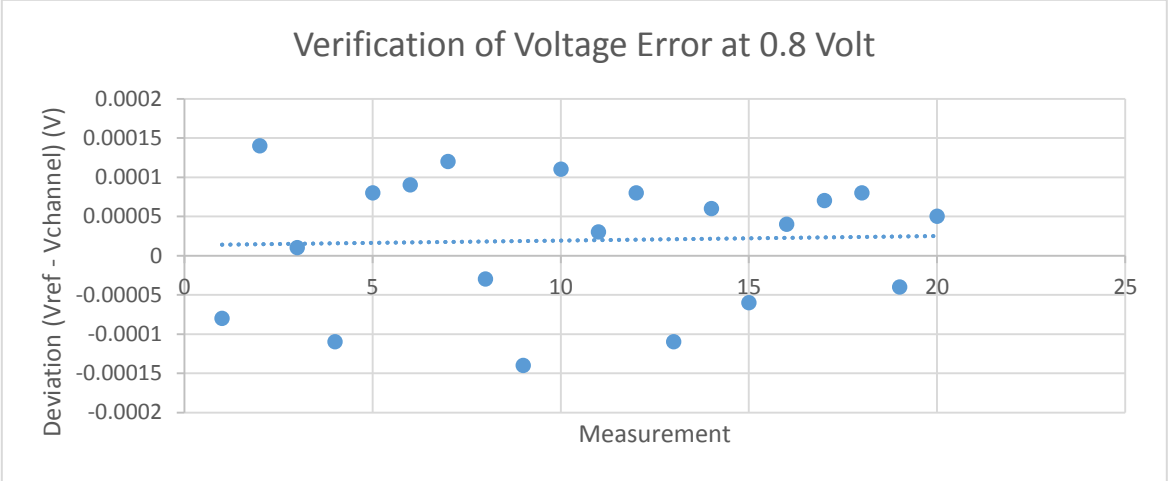


Figure 4.8. Voltage measurement distribution at 0.8 Volts

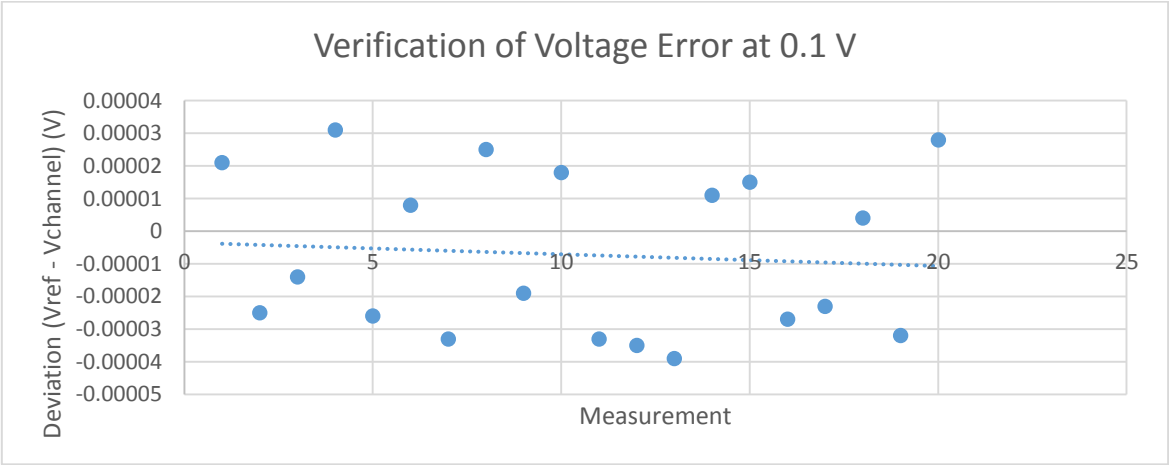


Figure 4.9. Voltage measurement distribution at 0.1 Volts

The uncertainty is statistically analysed from the measurement data series using the experimental standard deviation S.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}} \quad (84)$$

Where x_i is the reading and $\bar{x}$ is the mean of all the readings. In order to gain the best estimate of uncertainty for a normally distributed population, the experimental standard deviation of the mean is used.

$$U_1 = \frac{S}{\sqrt{n}} \quad (85)$$

where U_1 is the estimated standard uncertainty for the measurement set.

In order to gauge if the standard deviation is a small or large quantity when compared to the mean for data set the relative standard deviation (RSD) may be calculated from

$$RSD(\%) = \frac{S}{\bar{x}} \times 100 \quad (86)$$

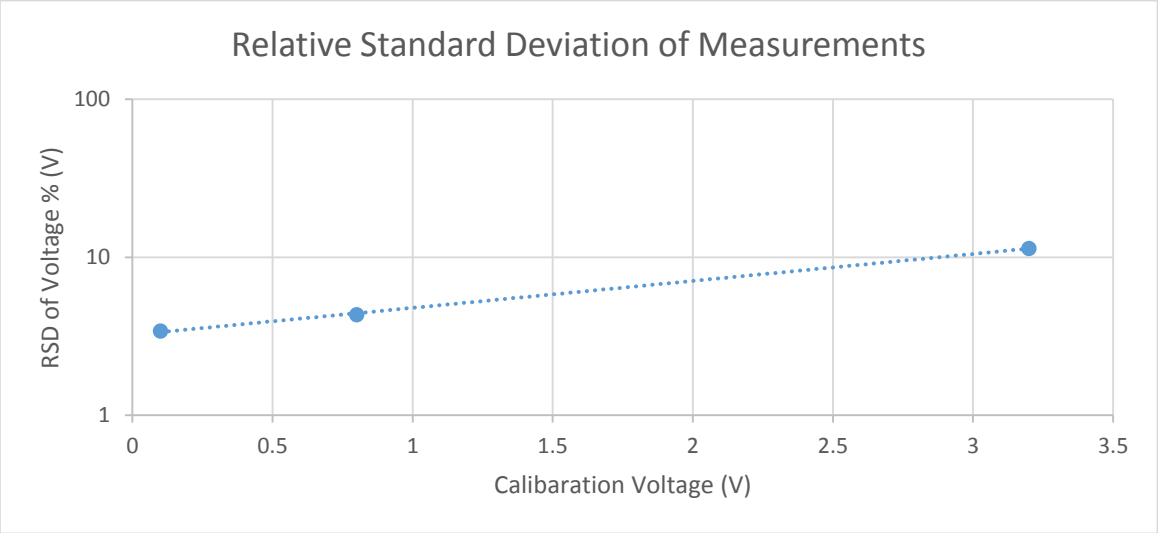


Figure 4.10. Relative standard deviation of measurements

To complete the analysis, numerous other considerations have to be taken into account. The first of these being how the data is distributed, for example, is the data Gaussian, Uniformly or Triangularly distributed. The degree of freedom, which is essentially a value relating to the amount of information that was employed in making the estimate. This is usually taken for type A uncertainties as being the sample size minus one (n-1), but is often considered infinite for parameters such as the manufactures specification. Lastly, the sensitivity coefficient is defined as the change in measurement response divided by the corresponding change in stimulus and defines how influential the measurement is.

In this case, from the Q-Q plots of the data (see section 4.4.2), the distribution is largely Gaussian in nature. This is a reasonable assumption, as most uncertainties, given a

large enough sample size from a population with a finite variance; as the mean of all samples from the same population will be approximately equal to the mean of the population, this being the Central Limit Theorem.

For the sample size, this was simply equal to the number of samples taken for each measurement, in this case 20. This sensitivity is in this case taken to be 1.

For each calibration point, the Type A estimated standard uncertainty is calculated using equation (85).

Calibration point (V)	Estimated Standard Uncertainty U_1
0.1	5.52×10^{-6}
0.8	18.9×10^{-6}
3.2	408×10^{-6}

Table 4.3. Showing the Estimated Standard Uncertainty for each calibration voltage

The Type B uncertainties need to be taken into account which cannot be determined statistically during the measurement process and include the error or inaccuracy of the calibrator, measurement error to the resolution of the NI9219 card, as well as lead effects, loading and thermal effects. A degree of judgement needs to take place and these estimates are expressed as one standard deviation for each of the different uncertainties.

The first Type B uncertainty due to the calibrator error U_2 estimates ± one standard deviation and is gained from the manufacturers recommended specification at the test point. For the voltage range being used, U_2 is ± 0.03%. The maximum instrument error per manufacturer's error at the point of test is (± 0.03 of the measured voltage). This measurement is based upon a normally distributed 99% confidence interval, hence to convert to ± one standard deviation,

$$U_{2std} = \frac{U_2}{2.58} \tag{87}$$

Lastly, the uncertainty pertained to the measurement limitation include error due to the resolution of the NI9219, this is perceived to be half of the least significant digit (LSD). To calculate the LSD for the NI9219, the measurement range divided by the number of bits used in the analogue to digital conversion, in this case 24 bits. The distribution is assumed to be rectangular in line with probability distributions for ADC in literature.

$$LSD = \frac{V_{ref}(+) - V_{ref}(-)}{2^N} \tag{88}$$

The standard uncertainty related to one LSD is then given by

$$U_3 = \frac{0.5 \times LSD}{\sqrt{3}} \quad (89)$$

Combining the uncertainties U_1, U_2 and U_3 with equation (82) gives the overall standard combined uncertainty U_c for the measurements to one standard deviation.

Tables (4.4, 4.5 & 4.6) below summarise the stages in calculating the uncertainties.

Source of Uncertainty	Type	U_i	Uncertainty Value (v)	Sensitivity Coefficient	Probability Distribution	Coverage Factor	Standard Uncertainty	Degrees of Freedom
Repeatability	A	U_1	408×10^{-6}	1	Normal	1	408×10^{-6}	19
Calibrator	B	U_2	960×10^{-6}	1	Normal	2.58	372×10^{-6}	∞
Resolution	B	U_3	381.5×10^{-9}	1	Rectangular	$\sqrt{3}$	110×10^{-9}	∞
Voltage Measurement	Combined	U_c	-	-	Assumed Normal	-	780×10^{-6}	253.8
Voltage Measurement	Expanded	U_m	1.53×10^{-3}	-	Assumed Normal	1.96	-	253.8

Table 4.4. showing the uncertainty analysis for the 3.2V measurements

Source of Uncertainty	Type	U_i	Uncertainty Value (v)	Sensitivity Coefficient	Probability Distribution	Coverage Factor	Standard Uncertainty	Degrees of Freedom
Repeatability	A	U_1	18.9×10^{-6}	1	Normal	1	18.9×10^{-6}	19
Calibrator	B	U_2	240×10^{-6}	1	Normal	2.58	93×10^{-6}	∞
Resolution	B	U_3	95.4×10^{-9}	1	Rectangular	$\sqrt{3}$	55.1×10^{-9}	∞
Voltage Measurement	Combined	U_c	-	-	Assumed Normal	-	112×10^{-6}	23423
Voltage Measurement	Expanded	U_m	219.5×10^{-6}	-	Assumed Normal	1.96	-	23423

Table 4.5. showing the uncertainty analysis for the 0.8V measurements

Source of Uncertainty	Type	U_i	Uncertainty Value (v)	Sensitivity Coefficient	Probability Distribution	Coverage Factor	Standard Uncertainty	Degrees of Freedom
Repeatability	A	U_1	5.52×10^{-6}	1	Normal	1	5.52×10^{-6}	19
Calibrator	B	U_2	30×10^{-6}	1	Normal	2.58	11.6×10^{-6}	∞
Resolution	B	U_3	11.92×10^{-9}	1	Rectangular	$\sqrt{3}$	6.9×10^{-9}	∞
Voltage Measurement	Combined	U_c	-	-	Assumed Normal	-	17.1×10^{-6}	1750
Voltage Measurement	Expanded	U_m	33.6×10^{-6}	-	Assumed Normal	1.96	-	1750

Table 4.6. showing the uncertainty analysis for the 0.1V measurements

The overall degree of freedom v_{eff} for the combined uncertainty has been calculated from the Welch-Satterthwaite formula, which considers each uncertainty U_i, each

sensitivity coefficient C_i and each uncertainty's specific value v_i for degrees of freedom to calculate v_{eff}

$$v_{eff} = \frac{U_c^4(y)}{\left(\sum_{i=1}^{N} \frac{C_i^4 U_i^4(x_i)}{v_i}\right)} = \frac{(780 \times 10^{-6})^4}{\frac{1 \times (408 \times 10^{-6})^4}{19} + \frac{1(372 \times 10^{-6})^4}{\infty} + \frac{1(110 \times 10^{-9})^4}{\infty}} \quad (90)$$

$$v_{eff} = 253.8$$

The expanded uncertainty can be calculated to give an overall confidence with the measurement result by expressing the uncertainty from a single standard deviation. This overall uncertainty is given by equation (83), repeated here for convenience.

$$U_m = kU_c$$

Where k is coverage factor which may be adjusted by the use of students' t distribution tables if the degree of freedom is small. Usually, a coverage factor of 1.96 is used, pertaining to a 95% coverage. Hence the overall $U_m = 1.96 \times 780 \times 10^{-6} = 1.53\, mV$ Finally, a measurement value including the quantified measurement uncertainties may be given for the NI9219. The voltage measurement will be $V = V_{avg} \pm U_m$.

Hence, for example, at 3.8 V, the maximum range of the calibration, the measurement will be $3.8 \pm 1.53\, mV$ at a confidence level of 95%.

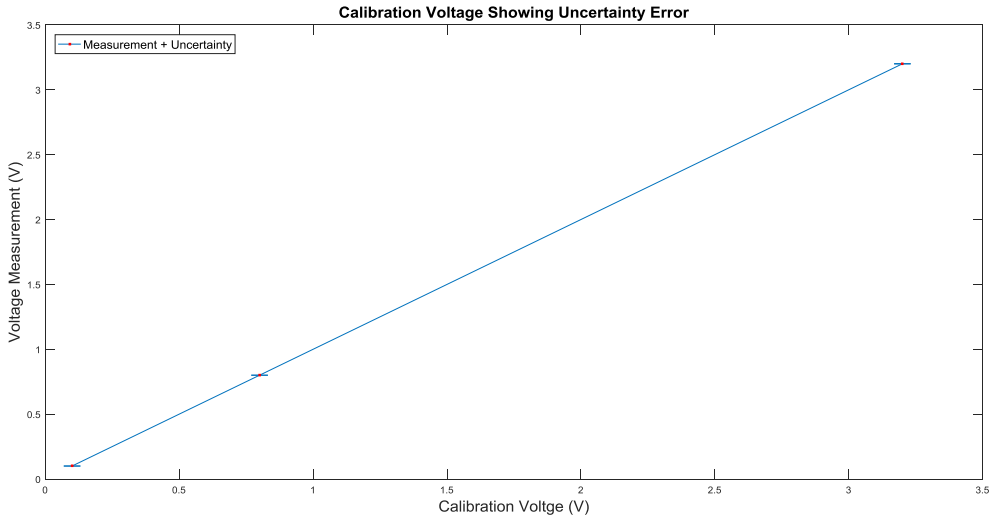


Figure 4.11. Voltage measurement against calibration voltage showing the error bars depicting the uncertainty error

4.4.2 Uncertainty within the Measured Data Analysis

In order to quantify the type of noise within the data an initial estimate may be gained from the statistical properties. The noise perturbed on the discrete measurement sequence forms a random distribution. Hence in a similar context, by treating the output contact resistance as a discrete sequence with noise from various sources superimposed upon it, the noise/uncertainty may be analysed and correlated to certain statistical distributions.

Probability distributions can be used to infer a degree of insight to the behaviour and nature of the random variables in the measurement sequence. A Q-Q (Quantile-Quantile) similar to P-P (Probability-Probability) plot provides graphical “goodness of fit” of the distribution of the data. By definition, a quantile is a class of variate that divides all the samples of a distribution into equal proportions. Generally, a normal Q-Q plot is widely used to estimate normality of the data. However, various other theoretical distributions can also be used to predict the distribution of the data set. In this case, the Q-Q plot compares data quantiles of the measured data set to quantiles of a known standard theoretical distribution.

From the Q-Q plots for each data set in figure (4.12), it can be seen that there is a strong Normal and Uniform component within the data sets. The process of quantisation is often assumed to be uniformly distributed, whereas thermal noise is assumed to be a normally distributed. For the sum of two or more independent and identically-distributed random noise variables they would lie more or less upon on each other, in this case there is a slight offset between each distribution.

Where the data points fall along a line in the middle of the graph, but curve off in the extremities the data exhibits more extreme values that would be expected if they truly came from a Normal or Uniform distribution. Hence, there is another underlying noise distribution present, which may be attributed to the process itself.

Where the Q-Q plot gives a useful insight into the nature of the distribution of the data, it lacks a rigorous quantitative measure of the type of uncertainty in the data. More analysis needs to be carried out to quantify the nature of the noise sources.

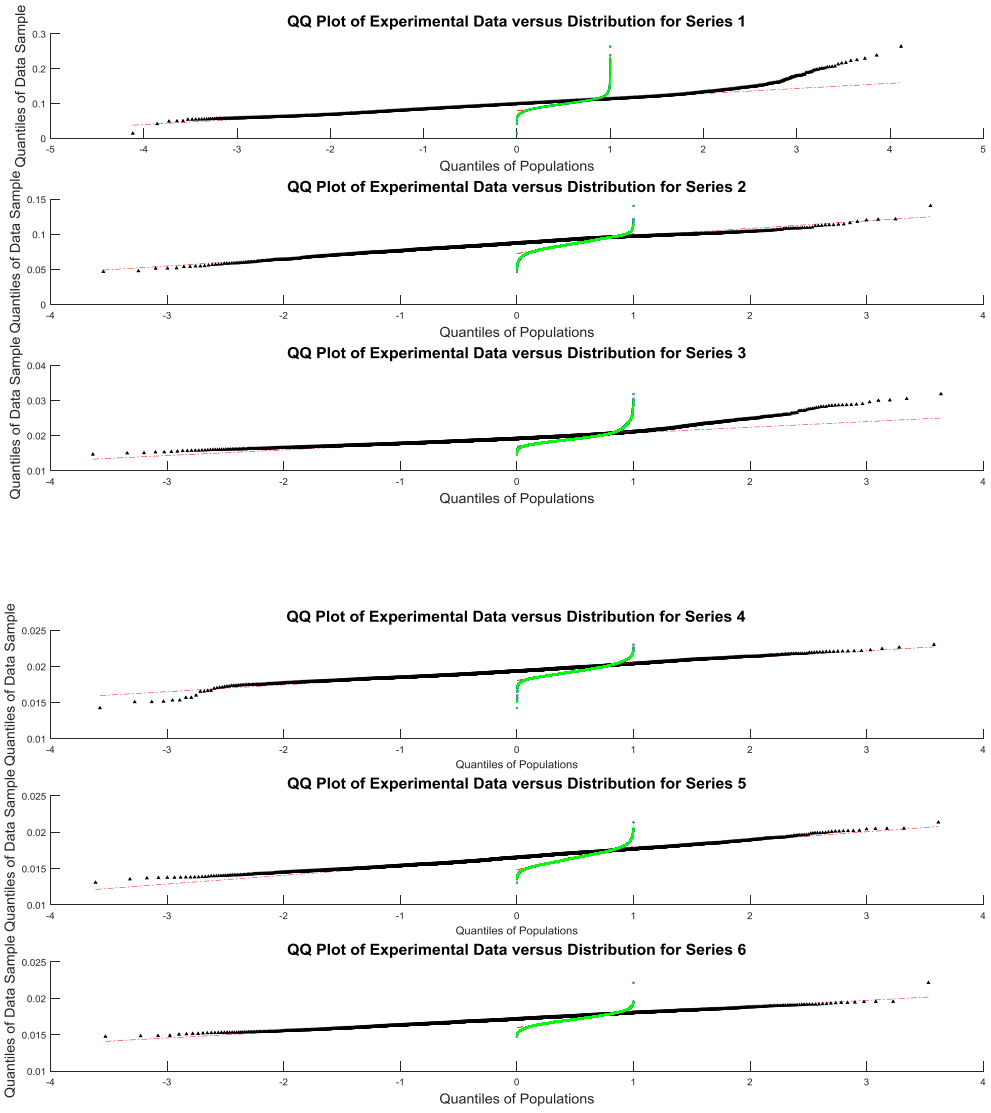


Figure 4.12. Q-Q plots of measured contact resistance (Green represents Uniform distribution, Black represents Normal distribution and the dashed red line represents the theoretical distribution.)

4.4.2.1 Noise Spectral Density

In order to try and gain a further insight into the noise within the measured data a spectral density analysis was carried out. The resulting spectral analysis occurred in the 0.1 V calibration range, as the signal to noise ratio is the lowest in this range. Noise spectral density is the average noise power per unit of bandwidth (dB/Hz) and is extracted by the use of a FFT on single measured data set. In order to eliminate biasing in the results, which may occur at lower frequencies due to the noise not being constant/flat, the root mean square (RMS) of the individual spectral densities is

calculated. The total noise spectral density may then be expressed over N acquisitions in terms of

$$\sigma_{v_o} = \sqrt{\frac{1}{n}\sum_{n=1}^{N}\left(\sigma_{v_o n}\right)^2} \tag{91}$$

and is expressed in $dBV/\sqrt{Hz}$.

From the results in figure (4.13) below, two inferences may be made. Firstly, the majority of the power spectrum is focused in the 100 Hz – 1000 Hz range and is typically flat with average noise spectral density concurring with the values of the variance of the measurement noise taken at across the measurement range. This is likely to be due to quantisation noise Therefore, the assumption is the noise is Gaussian in distribution throughout this range, backing up the assumption from the Q-Q plots.

Secondly, from 10 Hz there is a linearly decreasing flat noise source with a corner frequency of approximately 100 Hz, this from literature implies flicker noise. Flicker noise has a 1/f power spectral density and is often referred to as 'pink noise', it occurs in almost all electronic devices. In voltage measurements the 1/f noise is related to a direct current, the fluctuations in resistance are transformed into voltage fluctuations, this may also occur in resistances with no DC passing through them, and is a function of the temperature fluctuation modulating the resistance (Voss, Clarke, 1976).

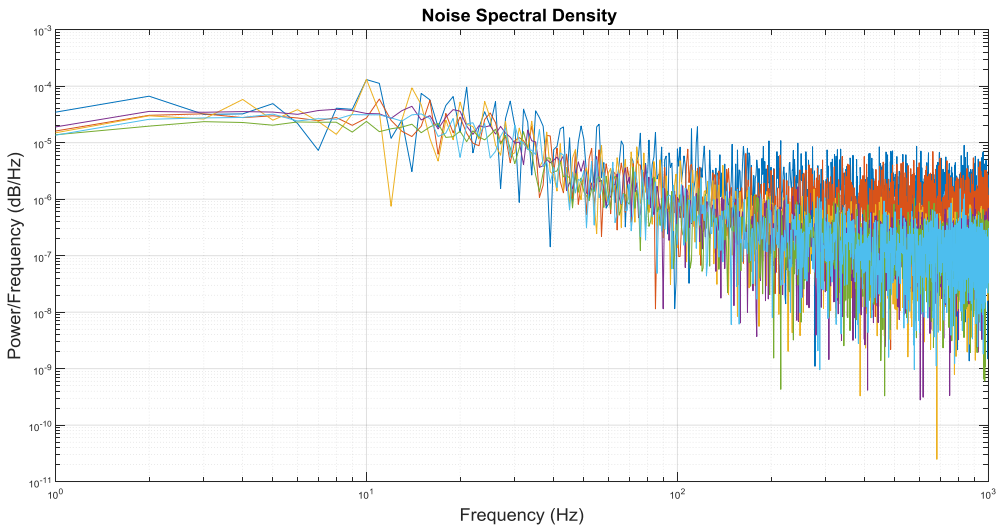


Figure 4.13. Noise spectral density for the experimental data

4.4.3 Theoretical Calculation of Noise in the Measurement

In order to further quantify the hypothesis that the noise in the measurement is due mainly to the quantisation, a theoretical analysis is undertaken.

4.4.3.1 Quantisation Noise Analysis

When digitising an analogue an ideal N-bit quantiser introduces a maximum error of $\pm\frac{1}{2}LSB$ (where LSB is the least significant bit), 1 LSB or quantisation level Δ is defined as

$$\Delta = \frac{V_{max} - V_{min}}{2^{N-1}} \tag{92}$$

where $V_{max} - V_{min}$ represents the full-scale amplitude of the analogue signal.

The non-linear quantisation noise due to rounding has a zero mean with average noise power (Vaseghi, 2008)

$$\mu_{QR} = 0\,;\ \sigma_{QR}^2 = \frac{\Delta^2}{12} \tag{93}$$

The truncation error has a non-zero mean with average noise power described by

$$\mu_{QT} = \frac{\Delta}{2}\,;\quad \sigma_{QT}^2 = \frac{\Delta^2}{3} \tag{94}$$

If it is assumed that the total quantisation noise is due only to rounding, which is valid for most ADC, the quantisation noise can be modelled as an additive, signal-independent, white noise uniformly distributed over the interval $[-\frac{\Delta}{2}, +\frac{\Delta}{2}]$ with variance defined (Proakis, 1996) by

$$\sigma_T^2 = \frac{\Delta^2}{12} \tag{95}$$

where σ_T^2 is the variance of the quantisation noise. The power spectral density of the quantisation noise over the flat and uniformly distributed region can be defined as

$$S(f) = \frac{\sigma_T^2}{\left({F_s}/{2}\right)} = \frac{\Delta^2}{6\,F_s} \tag{96}$$

Hence for the spectrum above, the PSD can be calculated as 9.93×10^{-7} dB/Hz.

4.4.3.2 Flicker Noise Model

The flicker power function $S_{nf}(f)$ (in W/Hz) is inversely proportional to the frequency

$$S_{nf}(f) = \frac{k}{f} \tag{97}$$

The constant k is arbitrary at this point.

The mean square voltage and current are proportional to their corresponding power, that is $V_n(f)^2 = S_n(f)R$, then

$$V_{fn}^2 = \frac{kR}{f} \tag{98}$$

It is common to relate flicker noise to the characteristic white noise at a frequency where both are equal called the noise corner frequency, f_{nc}

$$k = \frac{V_{wn}^2(f) \times f_{nc}}{R} \tag{99}$$

equation (98) now becomes

$$V_{fn}^2(f) = \frac{V_{wn}^2(f) \times f_{nc}}{f} \tag{100}$$

$V_{wn}^2(f)$here represents the Gaussian noise specified V/√Hz.

Integrating the 1/f function through the range of frequencies representing bandwidth gives the corresponding total mean square flicker noise

$$V_{fn-total}^2(f_1, f_2) = \int_{f_1}^{f_2} \frac{V_{wn}^2(f) \times f_{nc}}{f} df \tag{101}$$

Integrating, gives the following result

$$V_{fn-total}^2(f_1, f_2) = V_{wn}^2(f) \times f_{nc} \times \ln(\frac{f_2}{f_1}) \tag{102}$$

$$V_{fn-total}(f_1, f_2) = \sqrt{V_{wn}^2(f) \times f_{nc} \times \ln(\frac{f_2}{f_1})} \tag{103}$$

Using the noise approximation for the Gaussian white noise, and for a bandwidth of 90 Hz and a corner frequency of 100 Hz the flicker noise was approximated to be 14.9 μV.

The total noise power in the measurement therefore may be approximated by the combination of the quantisation noise and the flicker noise

$$\sigma_{measurement} = \sqrt{\sigma^2_{(\text{Quantisation})} + \sigma^2_{(\text{Flicker})}} \quad (104)$$

The noise overall was approximated to be 15.1 μV, which equates with the standard uncertainty measured at 0.1 Volts in table.

4.4 Median Filtering

A median filter is a non-linear class of filter and is particularly good for removing impulsive type noise from a signal (Qiu, 1994). The non-linear function of the median filter can be expressed as

$$y(n) = \text{med}[u(n-k), u(n-k+1), \ldots, u(n), \ldots, u(n+k-1), u(n+k)]$$

where $y(n)$ and $u(n)$ are the output and input signals respectively. The filter "collects" a window containing $N = 2(k+1)$ samples of the input signal and then performs the median operator on this set of samples. An important property of median filters is that they preserve edges, both positive and negative or stepwise discontinuities in the signal. The filtering results are shown in figure 4.14 below, showing the original noisy signal and the median filtered signal which replicates the original.

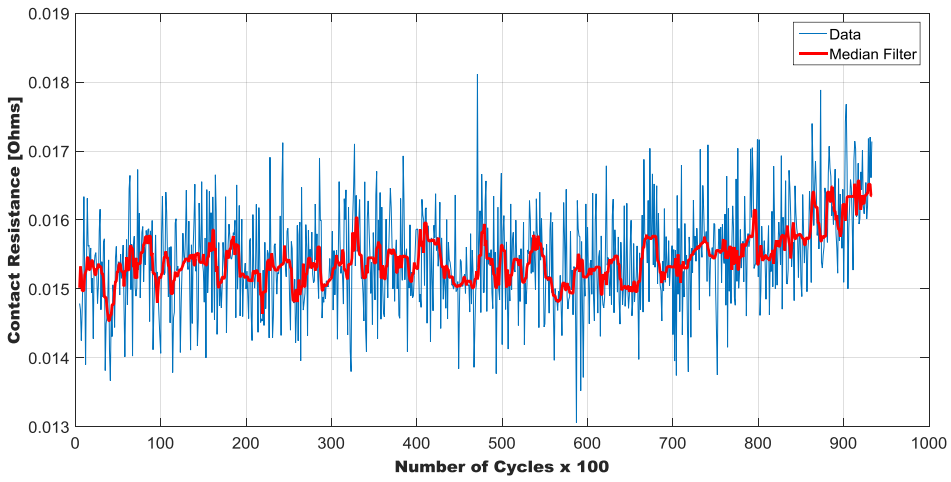


Figure 4.14. Median filtering of data

The filtered tests results of seven samples of contact resistance are shown below. The x-axis measurements are shown as number of cycles ×100, as the measurement was taken every 100 cycles.

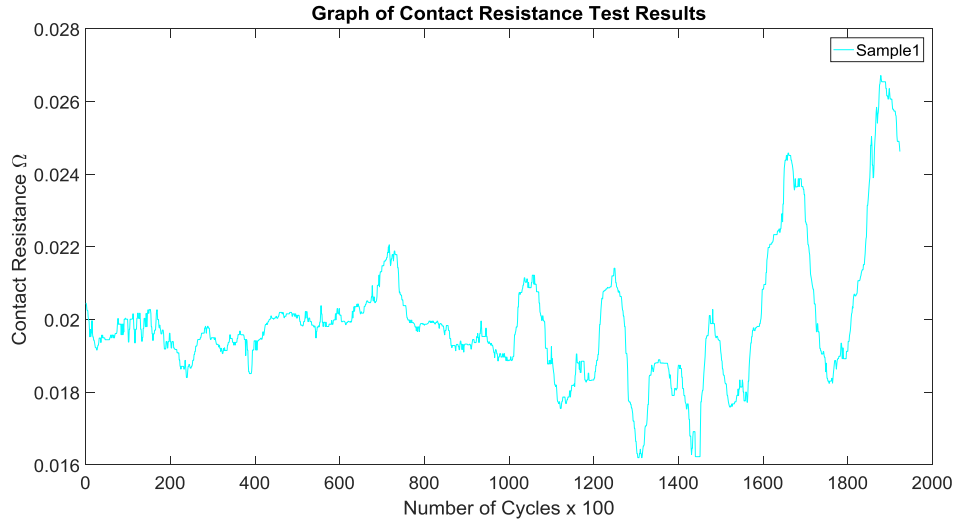


Figure 4.15. Sample 1

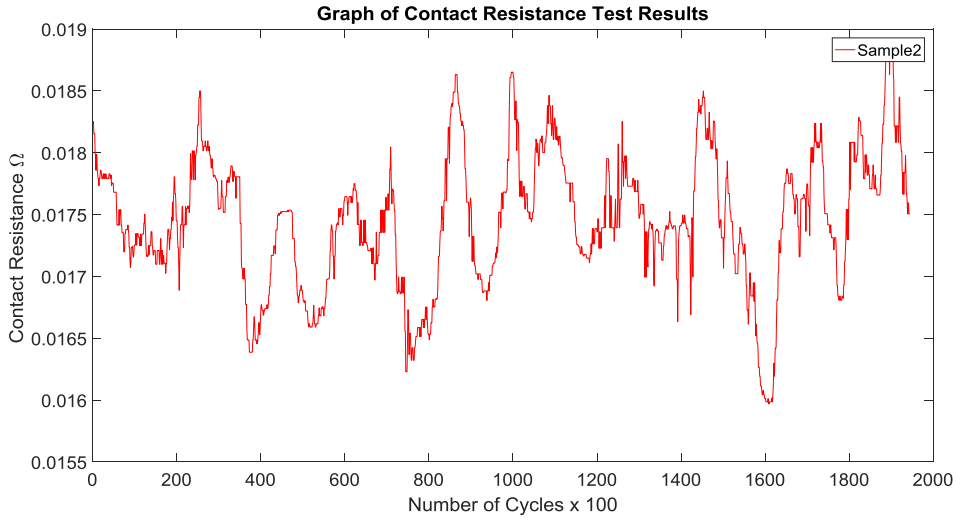


Figure 4.16. Sample 2

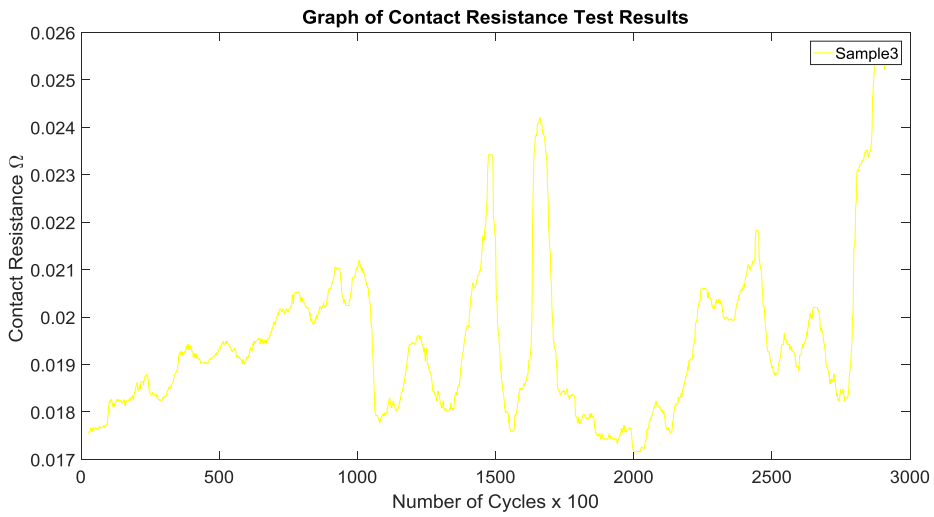


Figure 4.17. Sample 3

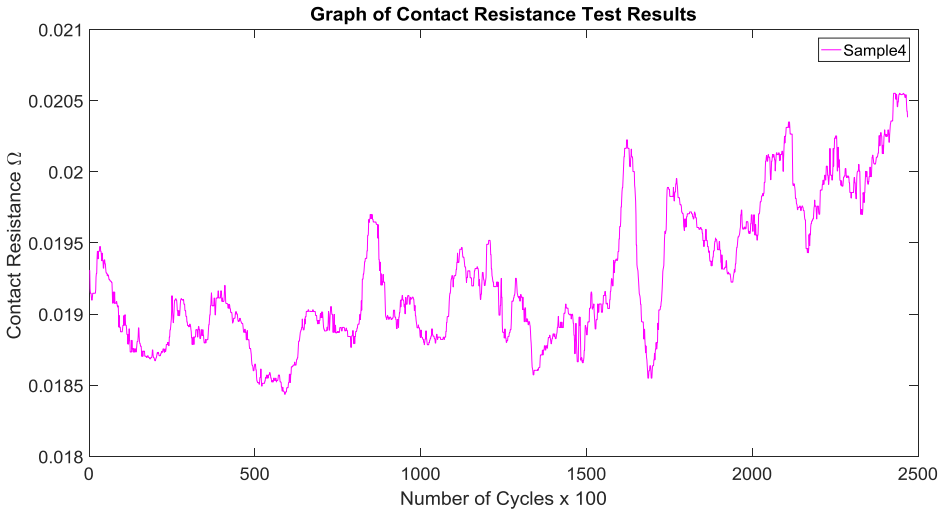


Figure 4.18. Sample 4

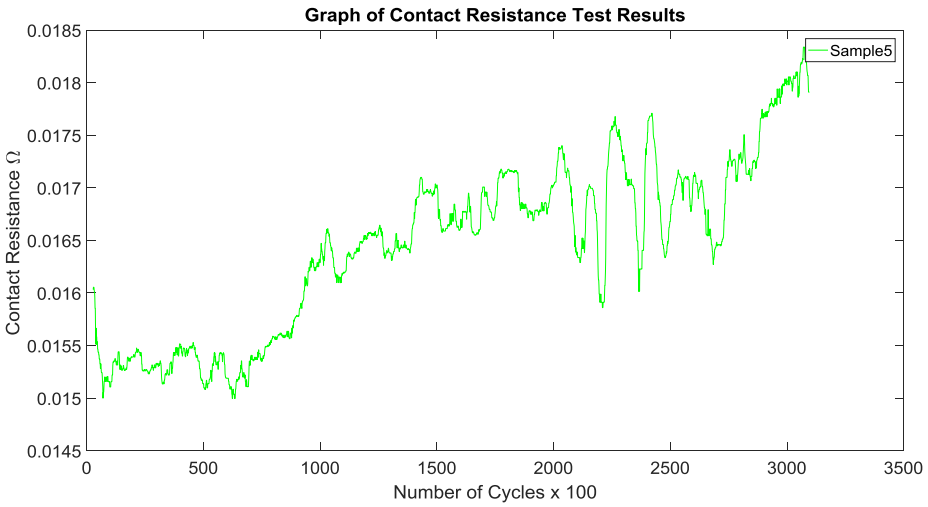


Figure 4.19. Sample 5

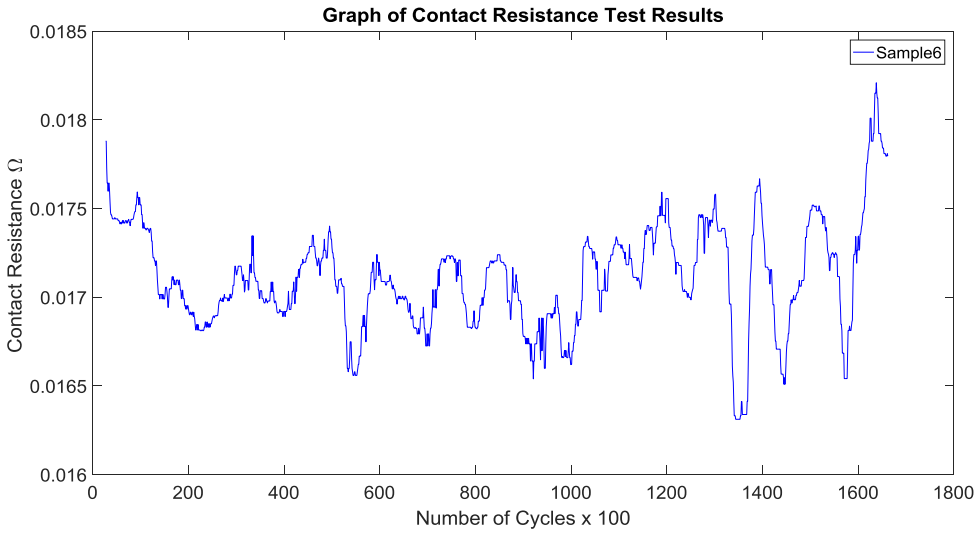


Figure 4.20. Sample 6

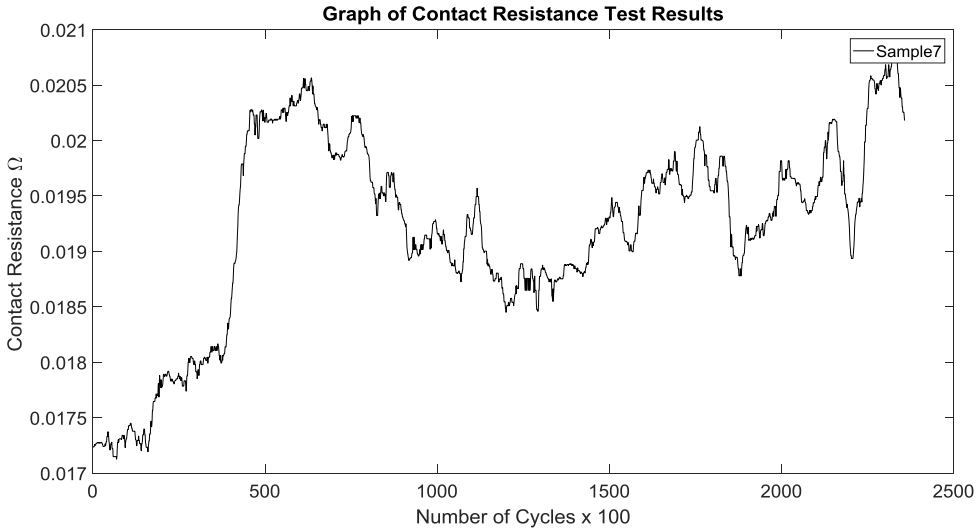


Figure 4.21. Sample 7

4.5 Test Results

The results show a general structure and certain boundaries may be categorised in Figure 4.22, showing the ensemble of test results. Firstly at **a**, for high voltage and currents (in excess of 1 A), the initial contact resistance will in most cases be high. This is due in part to the build-up of oxides and deposits on the surface of the contacts during manufacture and within storage/non-operation. This initial high resistance is quickly reduced by the effect of electrical cleaning due to fretting and the thermal destruction of the oxide layers, which sees the contact resistance drop back to a normal range. Hence manufacturers often give a value of maximum contact resistance

when the device is new, this is often way in excess to the normal operating contact resistance until the device has bedded in.

Section **b** depicts the climb in contact resistance up to the threshold of 100,000 cycles, this is the minimum number of cycles the manufacturer states contacts will operate for under a resistive loading. In general, a steady climb in the contact resistance may be observed up to this point. The effects of wear due to the physical effects discussed in chapter 2 become more apparent as the cycles increase.

The contact it makes is subject to bounce causing multiple interruption of the current which leads to erosion on the contact surface.

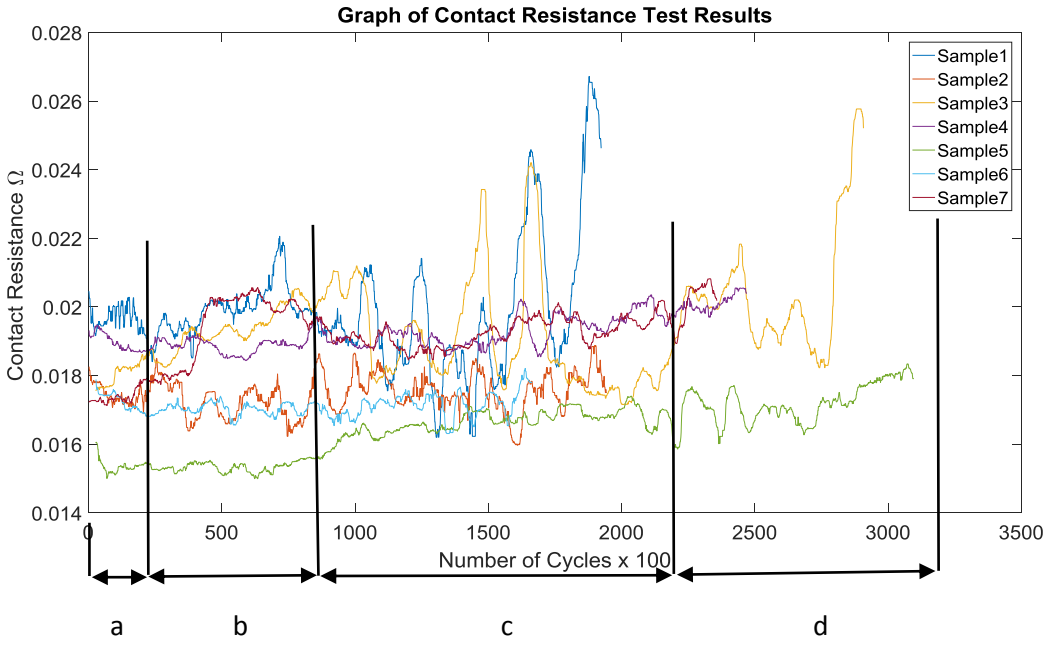


Figure 4.22. Montage of results, showing the various identified sections of contact degradation.

This severity of the contact erosion and how much material is transferred is dependent upon the load, current level, voltage, contact material, surface condition and bounce characteristics such as the gap, frequency, duration and timing. When this bounce subsides and tends to zero, the contacts close onto a molten liquid spot and weld together (known as dynamic welding).

The second part of the effects occur in section **c** when the contacts breaks. The moving contact opens to interrupt the current flow and when the contacts separate a sequence of two events occur on the surface contact. High current density heating of localized spots, spot melting, metal bridge transfer, formation of metal vapour arc and gaseous arc, arc transfer of contact material and finally arc extinction. This break

erosion depends on device characterisation, arc time and arc voltage and contact material, (Leung and Lee, 1991).

Section **b** & **c** therefore is made up of the superposition of both effects and gradual degradation of the contact will occur. In section **c** the damage becomes a lot more apparent, with the surface damage causing greater changes in contact resistance. However the underlying trend, despite the oscillating peaks and troughs is an increase in general and to some extent this may be characterised as fitting a linear best fit.

In section **d**, the eventual failure of all the relays under test was due to the contacts no longer opening due to welding. The oscillations in the resistance are more pronounced, with the eventual failure happening in most cases as the contact resistance takes a downturn. Failure is sudden and total, with complete loss of functionality. It can be seen from the number of cycles however, this failure is in all cases at least twice that of the stated minimum specified by the manufactures.

This coincides with the processes described in literature. The arc melts the surface of the contact for a period of time and after that the process of solidification begins, this can cause changes in the surface topography. (Leung, 2006) shows the effects for the arc in atmospheric air for Ag–SnO_2 and Ag–SnO_2–InO_2.

For contacts opening in air, the various gases present can make the change in surface structure very complex, this is in part due to the reaction permutations between the arc, the contacts surfaces and the constitute chemicals in the air. This may lead to the formation of oxides, nitrides and carbonates. Additional complication may arise due to additional pollutants such as dust, oils and greases, sulphides and chlorides that are present.

This change of surface topology shows as a change in the contact resistance. The resistance may increase, decrease or remain the same depending on the contact material, arc characteristics, atmosphere and mechanics, such as force on the contacts and if there is enough pressure or sliding action to rupture the surface film. The above failure modes due to impurities may be reduced by using hermetically sealed relay cases, filled with an inert gas such as helium or argon.

Even with hermetically sealed contacts, erosion however is still present, depending on the arc varicosity on the surface, this erosion may reveal new, un-oxidized material on each operation, and this will cause the contact resistance to remain at a low value. As

soon as a film forms, the closed contact will exhibit an increased temperature, leading to additional film formation. The process is summarized in figure 4.23 below.
In the above results, this becomes particularly evident, where as one may expect a general exponential rise to the end of life, the reality is far from this and can be explained from the above discussion as well as the section by section discussion.

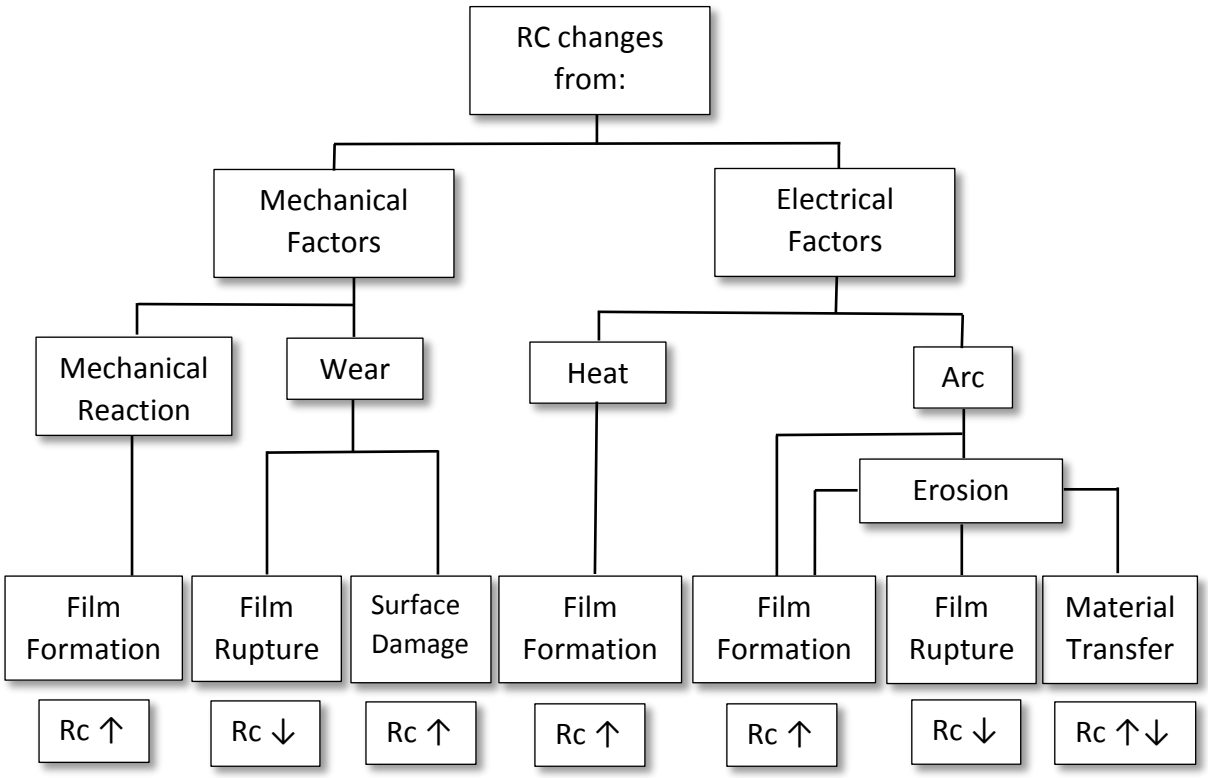


Figure 4.23. Summarizing the variation in Rc in an electrical contact.

4.6 Process Noise Effects on Contact Resistance Measurements

The factors governing the reliable operation of any electrical contact may be based on a great deal of interrelated and independent parameters. These parameters can be sub-divided into two elementary groups: Those which are internally occurring and subsequently, those which are externally occurring.

The category of internal factors represents the influence of mechanical effects, such as contact loading, sliding velocity and reciprocation and characteristics of motion, are commonly grouped under the term ‘fretting’. As well as electrical based factors, such as the type and strength of current and the operating voltage.

External based factors are representative of environmental effects and are often uncontrollable; effects such as time-temperature variation, humidity, atmospheric pressure as well as contaminants from air board particles.

These factors affect the performance of the contact by changing the properties of the contact material and depositing surface films. This in turn leads to physical and chemical changes in the contact, resulting in the build-up of wear with a subsequent effect of degradation of the contact interface, increase in contact resistance and eventual failure.

Relating these and previously discussed effects from arcing allows the quantification of process and measurement noise in the experimental results which are essential later on in the prediction process.

4.6.1 Fretting

A material that is subjected to small oscillatory movements at the interface of contacting materials can suffer from accelerated surface damage, this process is known as ‘fretting’. A unified model to explain the fretting process remains unfounded, however, the effect is dependent on numerous factors and many theories have been proposed, each is plausible to the exclusion of any other. The main factors that contribute to fretting can be roughly categorised as a) contact conditions b) environmental conditions and c) the properties and behaviour of the contact material, respectively. This may be summarised in figure 4.24 below.

An absence in literature of the effects of failure due to fretting is not without reason. As fretting is a time related process, the effects only become noticeable after long periods of time due to the build-up in the contact zone of wear based debris. As well as this, the destruction of the contact zone due to arcing and melting can shield the observer from any recognisable effects, especially at early stages.

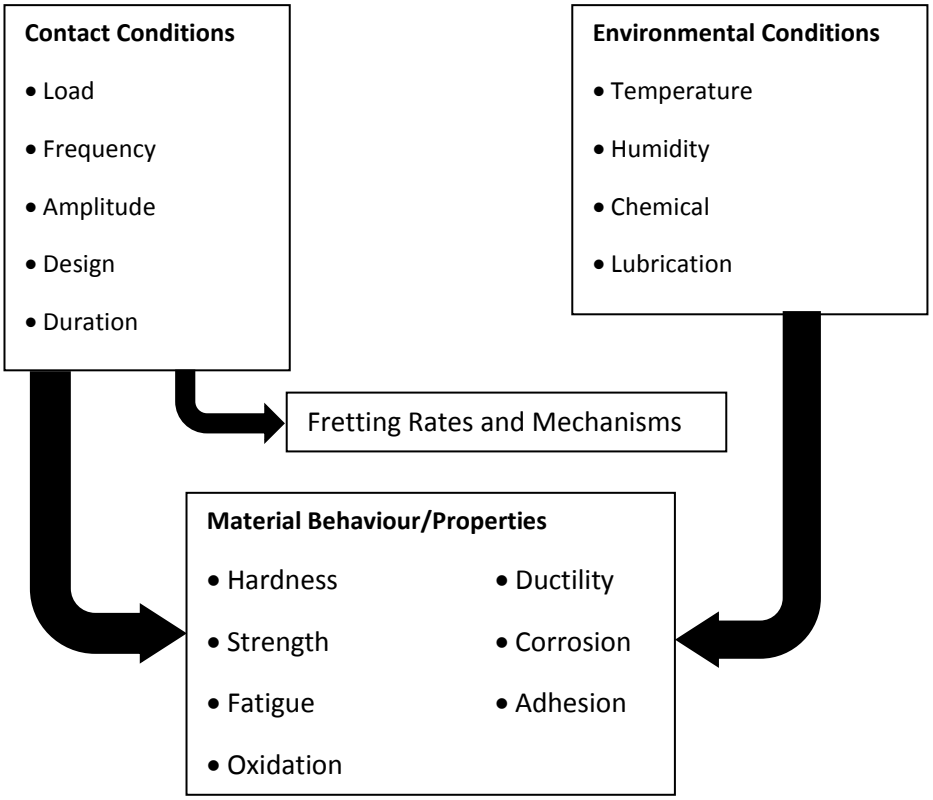
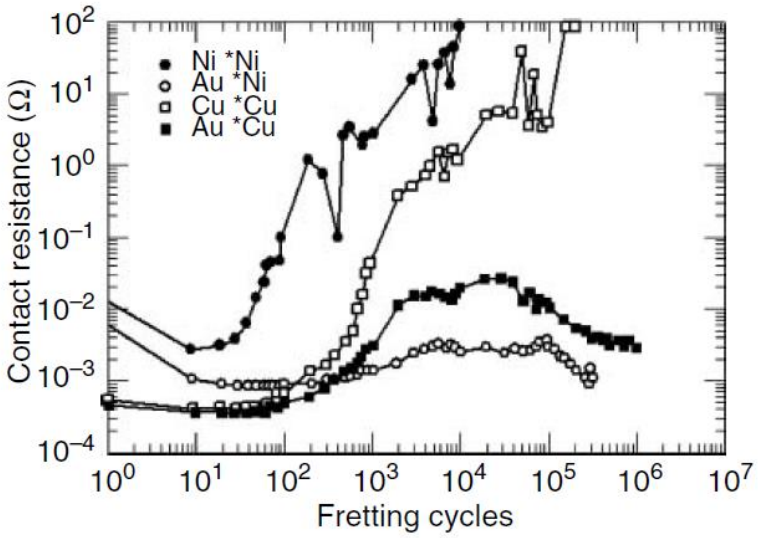


Figure 4.24 showing the effects contributing to process noise in measurements.

A few authors have carried out work into trying to measure the effects of fretting. The effects of fretting on the contact resistance of different contact plating materials and aluminum–tin-plated-copper and aluminum–copper was explored by Antler and Sproles, 1982; and Braunovic´, 1992. The results which are show in figure 4.25 a&b below.



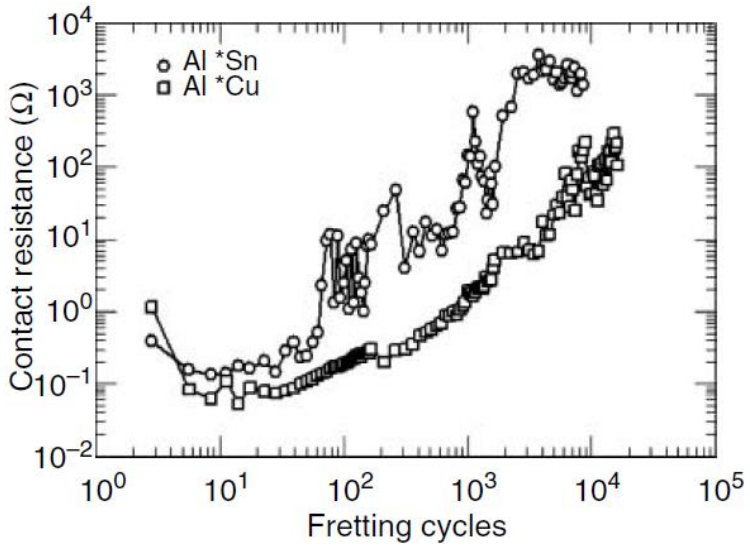


Figure 4.25 a&b, showing the effects of fretting on the contact resistance of different contact plating materials; aluminum–tin-plated-copper and aluminum–copper. Antler, M. and Sproles, E. S., IEEE Trans. CHMT, 5(1),158–166, and 1982; Braunovic´, M., IEEE Trans. CHMT, 15, 204–214, 1992.

The loading on the contact also exhibits considerable significance on the contact resistance due to fretting conditions. A small contact force can result in larger asperities in the measured contact resistance and even result in eventual open circuit. This result may be intuitive, as a greater loading results in less movement and hence less fretting.

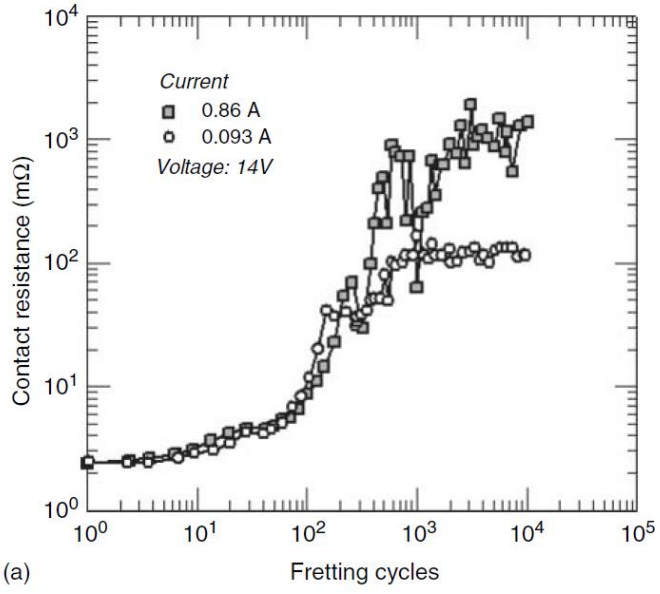
At loads less than 1 N, when contacts are made, the surface asperities of harder materials penetrate the oxide films naturally present on the material establishing localized metallic contacts and setting up conducting paths. Fretting causes the shearing of these metallic bridges which in turn causes the formation of wear products. Interestingly, a small fraction of these will oxidise, however, the majority will remain as metallic particles; thus a good metallic contact between the conducting surfaces is established. This effect is manifested by a **decrease** in contact resistance. This process is cyclic, and the temporary rupture of the insulating layer and appearance of localized metallic contacts and conductive paths is soon eradicated. Oxidation due to high current density will quickly eliminate the conducting paths resulting in a rapid **rise** in contact resistance. This effect is associated with low contact force only.

Hence contact force becomes important and subsequent decline within the components of the relay, such as the electromagnetic field in the coil and wear to associated mechanical components will mean force is also reduced further.

Other effects that can increase fretting are the frequency. The rate of fretting is shown to be dependent on the oscillation frequency. Oxidation will occur at lower frequency as the process is time dependant, hence, a decrease in the number of conduction paths and again an increase in contact resistance.
Environmental effects can cause changes in chemical reaction rates. Moisture can cause changes which affect the physical characteristics such as collection of debris and the surface mechanical properties of the contact material may be attributed to relative humidity. Related to this is the effect of temperature on the fretting process. The rate at which a chemical reaction takes place, such as oxidation and corrosion and the subsequent damage resulting, is in part, temperature governed.

4.6.2 Effect of Current

The effect of current across the contact interface has already been discussed in the literature review. Changes attributed to surface film formation and contact face asperities can cause localised heating and structural changes, resulting in changes in the contact resistance. Subsequently, the effect of current on the contact resistance behaviour of tin-plated copper contacts under fretting corrosion conditions was investigated in detail by (Lee and Mamrick, 1988).



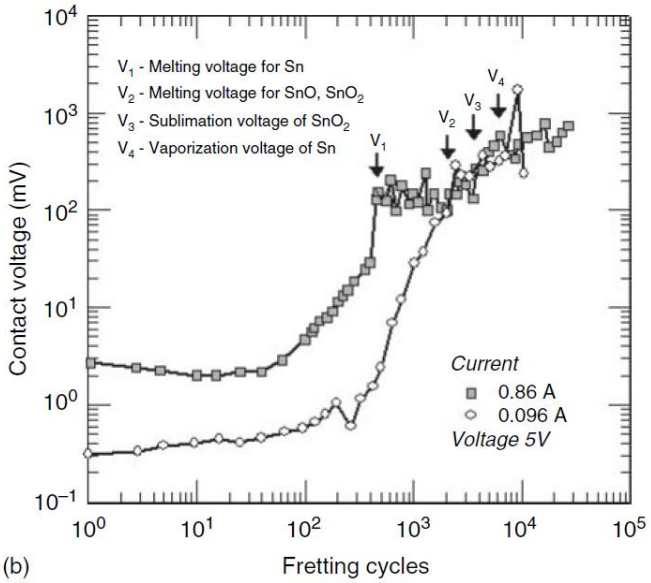


Figure 4.26 a&b. Lee, A., Mao, A., and Mamrick, M. S., Proceedings of 34th IEEE Holm Conference on Electrical Contacts, San Francisco, 87–91, 1988.

The results shown in Figure 4.26 a&b depict the contact resistance (a) and contact voltage (b) as a function of fretting cycles and electrical current. The conclusions drawn are that contact resistance behaviour can be explained by the presence of resistance plateaus that fluctuate delaying a further resistance rise. As the applied voltage and current are increased, it was found that the resistance plateaus become lower and longer.

To obtain a physical understanding of these resistant plateaus, the current through the contact constriction as discussed in (Chapter 2) causes the contact spot to thermally runaway until the melting of the material occurs (this constitutes the first plateau). Further damage to the contact in the form of corrosion gives higher resistance and more heating, the temperature can rise further to the melting, sublimation, and decomposition of the oxides, and even up to the vaporization of the material, collectively forming the second contact resistance or voltage plateau.

4.7 Threshold Determination

Investigation after testing deemed all failure was due to the contacts welding. Therefore, a threshold for failure was determined by calculating the theoretical contact resistance where a weld would occur given the type of contact material.

Contact welding can occur if a high enough current passes through closed contacts and causes the contact spot to melt. Welding can also occur after an arc is initiated

between contacts as they close. This arc can result from the electrical breakdown of the closing contact gap and it can also be continued by the contacts bouncing open once they have initially touched (Chen, 2010). Welding occurs when the contact area reaches the melting temperature T_m, then the voltage V_m across the closed contacts when they weld will be

$$V_m = (10^{-7}\{T_m^2 - T_0^2\})^{1/2} \quad (105)$$

Where T_0 is the ambient temperature. Equation (105) is a simplified version of the Wiedemann–Franz Law, which relates the ratio of the electronic contribution of the thermal conductivity (K) to the electrical conductivity (σ) of a metal, and is proportional to the temperature (T).

$$V_m = I_{weld} \times R_c \quad (106)$$

Where I_{weld}, is in amperes and R_c is the contact resistance in ohms. From the above, an estimation of the threshold where welding can occur may be calculated. Using equation () this may be written as

$$I_{weld} = \frac{1}{9\rho}\sqrt{\frac{F}{H}}\left\{(10^{-7}\{T_m^2 - T_0^2\})^{1/2}\right\} \quad (107)$$

Where the conductivity ρ is in Ω cm, Hardness H in kgf/mm^{-2} and F in Newtons. This equation for the weld current in terms of a function of temperature, force, hardness and conductivity, may then be used to give an approximation of the failure threshold by resubstituting and rearranging for R_c.

Conclusion

The analysis of the data poses several problems, firstly how and where to set the threshold of impending and total failure. This is problematic; where there is in general an upward trend in contact resistance and increased oscillations in all the results, there is however little correlation between them. This is despite the same controlled test conditions being applied, which illustrates the overall complexity of the combined failure process. An approximation to this has been proposed, however this is still quite abstract, with welding occurring when not expected.

The second problem is the choice of prognostic approach, which again is subject to difficulties again due to the failure pattern. The next two chapters will propose solutions to solve this problem by looking at a physics based and data driven solutions respectively and compare their respective effectiveness.

This chapter provides the following conclusions and contributions

- A method has been developed for the obtainment of real time measurements to assess the life of the relay using contact resistance.
- A data set has been produced that describes relay failure. This has been broken down into sections to enable the construction of a visual analytics model to enable to application of prognostic methods.
- The failure modes have been pictured as an early decision tool.
- The measurement noise within the data has been quantified.
- The process noise that leads to anomalies within the results has been explained.

Chapter 5 - Physics of Failure Damage Model Development for Relays

Introduction

From the literature review in Chapter 3, Physics Based Models (PbM) tend to provide the most accurate and precise solutions for prognostics, as well as providing confidence boundaries when projected with Particle and Kalman filtering, the dynamics of the states can also be estimated and predicted at each time interval. However it is also recognised that these type of models require a high level of knowledge of the component behaviour to form a complete model and often use experimental data to develop a model, which depends upon conditions such as environmental or material.

From the review of relay reliability in Chapter 2, there are essentially three constituent parts to the development of a model for the prognostics of relays; firstly, there are the components of the coil e.g. the windings and the associated structural parts such as the spring, sliding parts and insulation. Secondly, there are the contacts and the transfer of energy through them, this will cause heating due to Joule losses and arcing. Thirdly, there is the degradation of the contacts, which has been discussed in prior chapters. The primary failure mode due to transfer and wear of contact material is due to arcing, welding and bridging of the contacts, sticking, corrosion and foreign matter.

It has already been stated, that failure due to the contacts is many magnitudes more likely than that from the coil and associated components according to literature and manufactures test data. Therefore, concentrating on a prognostic for the contact seems a more proactive option, however, some metrics of wear may involve the use of measured parameters associated with the coil, e.g. pick up and release times, current through the coil, force on the contacts, arcing time and the contact position with time.

The making and breaking of the contacts incurs damage to the surface, primarily due to arcing. Thus a suitable model of the energy transferred by arcing into the contact is needed that reflects the damage from the excessive temperatures imposed by the arc. This energy transferred by the arc into the surface of the contact will cause a rapid

rise of the temperature of the contact that will be conducted in the form of heat and ultimately melting of the contact.

The losses of material incurred will be due to vaporization of the contact material, this may be viewed as being accumulative over time, causing a reduction in the affective contact area and change in the relative contact resistivity. This accumulative material loss from the surface due to the arc area will form the basis of the loss model, which will be in the form of a state equation and will be used for a prognostic prediction of RUL.

5.1 Development of Physics Based Model of Relay Contact Degradation

Initial work looked into the development of a physics based model to enable the degradation of the relay contacts to be estimated and the possibility for model based prognosis. A complete model would need to take into account the effects of contact breaking, bounce and travel position; thermal heating of the contact due to arcing and lastly material loss and the effect on contact resistance.

The proposed method is to construct an initial physical model of the relay body and use this to calculate the first three parameters, namely coil current, bounce and force, in which additional parameters may be estimated, such as arcing time, contact force, opening/closing time and contact position. The DC power from a simple arc model would then be used to estimate the power into the contact face. Thermal effects would be modelled by a heat equation and this would be used to estimate mass loss. The mass loss then is used in the contact resistance equation (4) outlined in Chapter 2 to estimate the RUL of the equation.

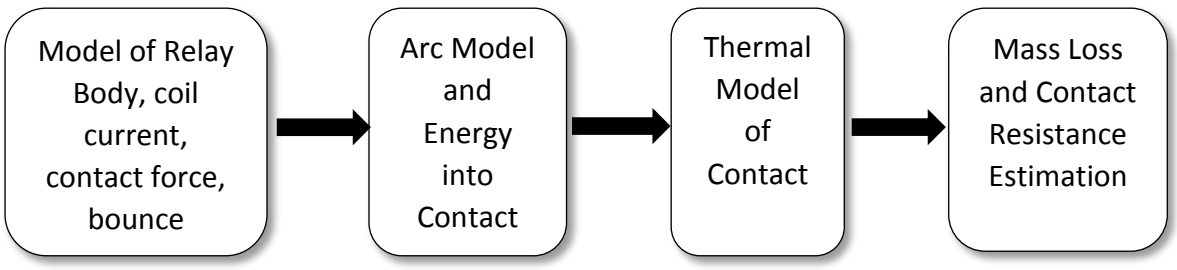


Figure 5.1. Block diagram showing each stage of model derivation

5.1.1 Derivation of a State-Space model for an Electromagnetic Relay

The Electromagnetic relay in its basic form can be broken down into a number of elementary components, namely; a coil formed from a number of turns of wire wound on a ferrite former with some associated resistance and inductance; a set of contacts which are connected via a pivoted lever with a return spring and a degree of dampening. This may be modelled as shown in figure 5.2 below where U represents the driving signal to the relay, R_{coil} is the resistance of the coil windings, F represents the electro-magnetic force in Newton's, k is the spring constant in N/m and b is the damping coefficient of the spring and pivoted arm. The distance x in metres represents the position the contacts from open to closed.

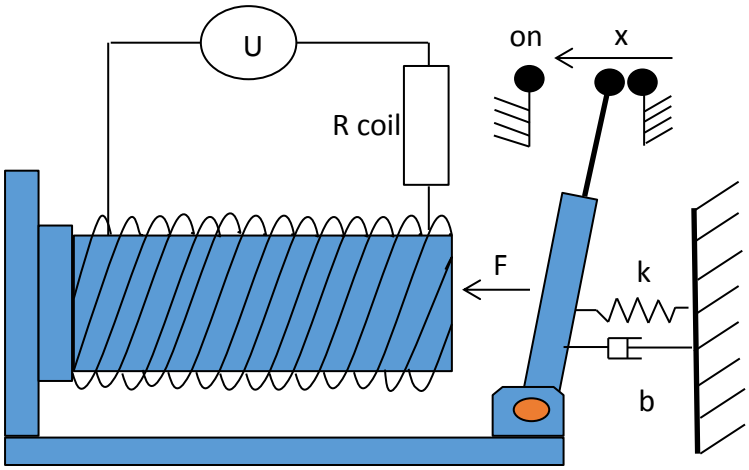


Figure 5.2. Depicting a simplified model of a relay

5.1.2 Lagrange Equations

In order to model such a system which contains a large number of lumped components which are both electrical and mechanical, the use of Lagrangian Mechanics (Dutton, Thompson and Barraclough, 1997) is used.

The Lagrange equations for determining the equations of motion of a dynamic system based on Hamilton principles may be stated as 'For a dynamic system in which the work of all forces is accounted for in the Lagrangian, an admissible motion between specific configurations of the system at the times t_1 and t_2 is a natural motion if, and only if, the energy of the system remains constant.'

This is applicable to both forces and motions as well as electrical systems in terms of voltages and currents. For a conservative system subject to external forces (neglecting energy losses), Lagranges equation may be written as:

$$\frac{d}{dx}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0 \quad (108)$$

Where $L = T - V$ is the Lagrangian (T and V are the kinetic and potential energy in the system) and q_i represents generalised coordinates.

For more general systems (that is, ones including power dissipation), Lagranges equation may be written as:

$$\frac{d}{dx}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + \frac{\partial P}{\partial \dot{q}_i} = Q_i \quad (109)$$

where P is the power function, describing the dissipation of energy by the system and Q_i are the generalised external forces acting on the system.

'The number of degrees of freedom of a body is the number of independent quantities that must be specified if the position of the body is to be specified (uniquely defined). Any unique set of such quantities is referred to as a set of generalised coordinated for the systems.'

Energy forms for linear mechanical and electrical elements can be represented as in table 5.1 below (Dutton, Thompson, and Barraclough, 1997).

Energy Type	**Mechanical**	**Electrical**
Kinetic Energy T	Mass M $T = \frac{1}{2}m\dot{x}^2$	Inductor L $T = \frac{1}{2}L\dot{q}^2$
Potential Energy V -	Spring K $V = \frac{1}{2}kx^2$ Gravitational M $V = mgh$	Capacitor C $V = \frac{1}{2}Cv^2 = \frac{1}{2C}q^2$ -
Dissipative Energy P	Damper b $P = \frac{1}{2}b\dot{x}^2$	Resistor R $P = \frac{1}{2}R\dot{q}^2$

Table 5.1 Energy forms for linear mechanical and electrical elements

note: current i is expressed in terms of charge q as $i = \frac{dq}{dt} = \dot{q}$ and linear velocity v is expressed in terms of displacement x as $v = \frac{dx}{dt} = \dot{x}$

5.1.3 Derivation of Relay State Equations

The Lagrangian equations discussed above may now be used to form a model for the electro-mechanical relay. Starting with the Lagrange equation for a non-conservative system, equation (109), $\frac{d}{dx}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + \frac{\partial P}{\partial \dot{q}_i} = Q_i$

The Lagrangian L may be written as: $L = T + W_e + W_m - V$

where T is the Kinetic energy, in this case due to the mass $T = \frac{1}{2}m\dot{x}^2$ and the electrical energy $W_e = 0$, and for the magnetic energy $W_m = \frac{1}{2}L(x)\dot{q}^2$. For the potential energy, in this case stored by the spring $V = \frac{1}{2}kx^2$, and also due to the mass $V = mgh$.

The power dissipation in the system is due to the damping of the arm and the resistance of the coil in the relay, $P = \frac{1}{2}b\dot{x}^2$ and $P = \frac{1}{2}R\dot{q}^2$ respectively.

Hence, substituting into equation (109) leads to the following equation for the relay.

$$L = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}L(x)\dot{q}^2 - \frac{1}{2}kx^2 - mgh \qquad (110)$$

The relay has two degrees of freedom e.g. independent quantities that must be specified if the position of the body is to be specified, namely due to the position of the contacts in terms of the displacement in the x plane and also due energy in terms of the magnetic force due to the coil. Hence these parameters need to be used for the quantities in the generalised coordinates for specifying the system.

Firstly, solving for the x coordinates, and making use of the Lagrange equation for a conservative system, equation (108) gives,

$$\frac{d}{dx}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0 \qquad (111)$$

Defining each differential gives:

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x}, \qquad \frac{\partial L}{\partial x} = \frac{1}{2}\frac{\partial L(x)}{\partial x}\dot{q}^2 + kx + mg, \qquad \frac{\partial P}{\partial \dot{x}} = b\dot{x}$$

Substituting into equation (108) above and rearranging gives

$$m\ddot{x} + b\dot{x} + kx + mg - \frac{1}{2}\frac{\partial L(x)}{\partial x}\dot{q}^2 = 0 \qquad (112)$$

secondly for the q coordinates we make use of the equation for a non-conservative system, equation (107).

$$\frac{d}{dx}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + \frac{\partial P}{\partial \dot{q}_i} = U$$

where U is the voltage applied to the relay coil.

The energies may be defined as

$$T = \frac{1}{2}L(x)\dot{q}^2, \qquad L = \frac{1}{2}L(x)\dot{q}^2 + \frac{1}{2}R\dot{q}^2$$

Differentiating gives

$$\frac{\partial L}{\partial q} = 0, \qquad \frac{\partial L}{\partial \dot{q}} = L(x)\dot{q}, \qquad \frac{\partial P}{\partial \dot{q}} = R\dot{q}$$

Substituting into equation (107) gives

$$L(x)\ddot{q} + R\dot{q} = U \qquad (113)$$

Substituting into equation (108) and (109) and making use of the fact $\dot{q} = \frac{dq}{dt} = i$ the equations may be written as:

$$m\ddot{x} + b\dot{x} + kx + mg - \frac{1}{2}\frac{dL(x)}{dx}i^2 = 0 \qquad (114)$$

$$L(x)\frac{di}{dt} + Ri = U \qquad (115)$$

The inductance of a coil $L(x)$ may be stated as

$$L(x) = \frac{N^2\mu_0 A}{x} \text{ and } \frac{dL(x)}{dx} = -\frac{N^2\mu_0 Ai}{x^2} \qquad (116)$$

where N is the number of turns on the coil, μ_0 is the permeability $4\pi \times 10^{-7}NA^{-2}$, i is the current in Amps and x is the length of the coil.

The flux linkage φ of a coil can be defined as: $\varphi = L(x)i$ and the magnetic energy stored in the coil is given by $W_m(\varphi, x) = \int_0^{\varphi} \frac{\varphi}{L(x)} \partial\varphi = \frac{\varphi^2}{2L(x)}$

The current in the coil can be defined as the rate of change of magnetic energy with respect to the flux linkage hence $i = \frac{\partial W_m}{\partial \varphi}$ and the force from the coil can be defined as the rate of change of magnetic energy over the distance the contact moves through.

$$F = \frac{\partial W_m}{\partial x} = -\frac{\varphi L(\dot{x})}{2L^2}$$

The rate of change of flux linkage can now be defined as

$$L(x)\frac{di}{dt} = \frac{\partial \varphi}{\partial t} - \frac{dL(X)}{dx} i \frac{dx}{dt} \quad (117)$$

We can now use the results (116) & (117) to redefine equations (114 and 115).

$$m\ddot{x} + b\dot{x} + kx + mg - \frac{\varphi^2 i}{2N^2 \mu_0 A} = 0 \quad (118)$$

Therefore equation (113) may be defined as

$$\frac{\partial \varphi}{\partial t} - \frac{\varphi}{2N^2 \mu_0 A}\frac{dx}{dt} + iR = U \quad (119)$$

putting $a = N^2 \mu_0 A$ and using $i = \frac{\varphi}{L(x)}$ and where $L(x) = \frac{a}{x}$ the above equations (118) and (119) can now be simplified.

The final equations are written as

$$\frac{\partial \varphi}{\partial t} - \frac{\varphi^2}{2a}\frac{dx}{dt} + \frac{\varphi x}{a} R = U \quad (120)$$

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx + \frac{\varphi^3}{2a^2} = 0 \quad (121)$$

as the component due to gravity has little or no effect

The above equations represent the electromagnetic and mechanical system and can be put into a state space format for easier computation.

Putting

$x_1 = \varphi,\ x_2 = x,\ x_3 = \dot{x}$ equations (120 and 121) can be written as

$$\dot{x}_1 = \frac{x_1^2 x_3}{2a} - \frac{x_1 x_2 R}{a} + U$$

$$\dot{x}_2 = x_3$$

$$\dot{x}_3 = -\frac{b}{m}x_3 - \frac{k}{m}x_2 - \frac{x_1^3 x_2}{2a^2 m} \quad (122)$$

The form of the state equations are non-linear and to assist analysis, need linearising around an operating point.

5.1.4 Linearisation Procedure and Results

Given a non-linear function $f(x_1, x_2, \dots, x_n, u_1, u_2, \dots, u_n)$ the operating condition need to be defined which will produce a linear model. This condition is often found by setting $f(x_1, x_2, \dots, x_n, u_1, u_2, \dots, u_n) = 0$ and considering the resulting values of $x_1, x_2, \dots, x_n$ for a set of input values $u_1, u_2, \dots, u_n$.

Partial differentiation is used to find the linear coefficients and evaluate these at the defined operating points. This may written in matrix form, firstly values for the various constants are substituted. Matlab™ was used to simulate the equations based on the following parameters

$\mu_0 = 4\pi \times 10^{-7} NA^{-2}$, N=1000 Area = 0.001 m^2 hence $a = N^2 \mu_0 A$

a = 7.9 x 10^{-4}, b =0.9, m = 0.01 kg, k = 10, R = 10 Ω

The graph in figure 5.3 shows the current through the coil which rises exponentially according to the equation $i = I_0(1 - e^{-\frac{Rt}{L}})$ where I_0, is the initial coil current, R is the resistance and L is the inductance of the coil, respectively.

Contact position, contact bounce and settling time is shown in figure 5.4. The initial position is shown at t = 0, where the contact first closes, here the contact face is displaced by 0.3 mm due to bouncing, and then over 50 mS, subsequent exponential decaying in the bouncing occurs until the contact faces come together with a displacement of 0 mm.

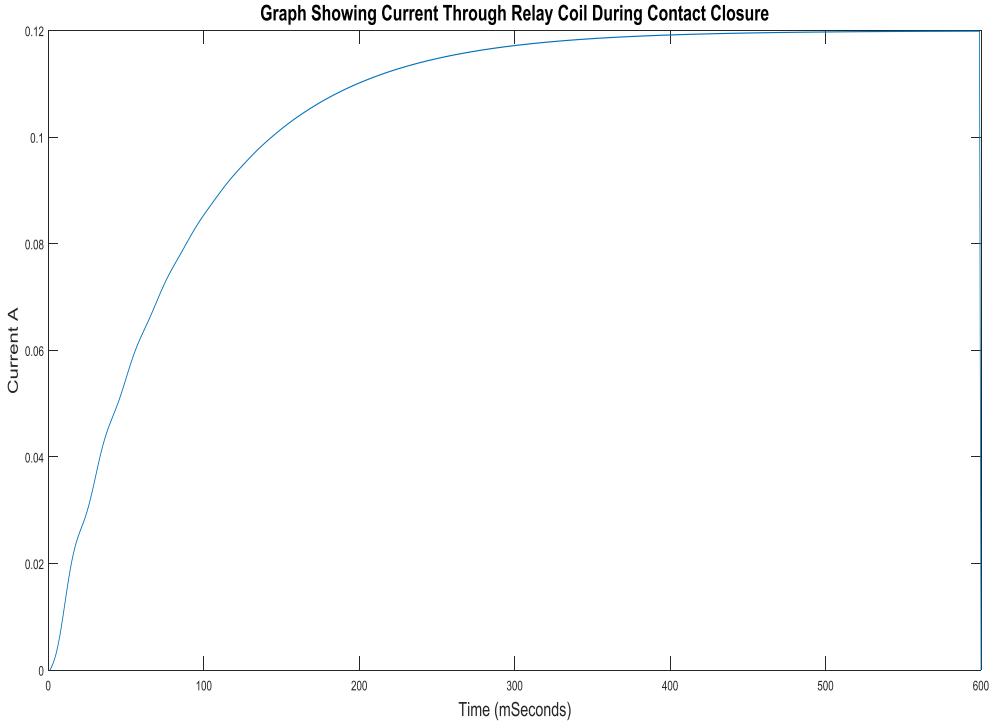


Figure 5.3. Showing simulated coil current with time

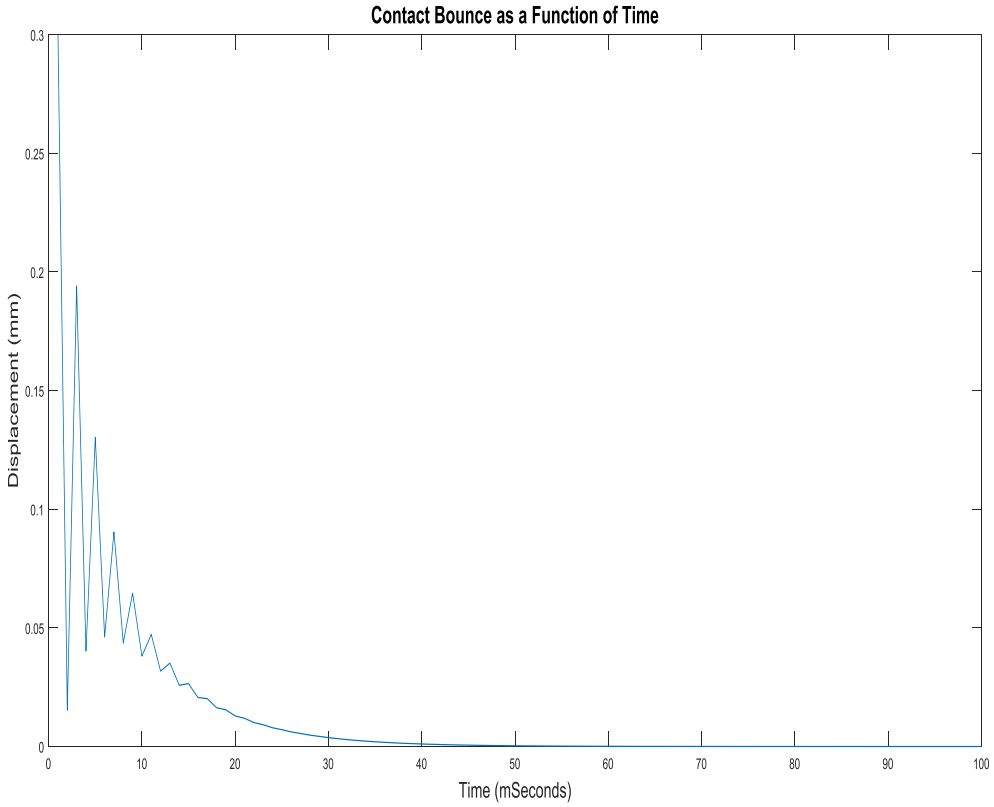


Figure 5.4. Showing simulated contact position with time including the bounce time.

5.2 Heat Model and Mass Loss Estimate

The second stage in the model development is to gain an idea of the energy dissipated through the contact due to arcing and the subsequent heating and resulting damage through material vaporization. The procedure follows the work done on mass loss by (Swingler and McBride, 1996), by developing a model of the arc energy transfer and the heat flow through the contact. From this, vaporization of material may be estimated and subsequent mass loss.

5.2.1 Calculation of Arc Energy

The physics of arcing is extremely complex and not fully understood. Over the years numerous models have been proposed in literature to enable the effect of arcing in power systems to be simulated and understood. A great deal of the early work revolved around defining the volt-ampere characteristics of the arc experimentally and was dependent upon the test conditions, gap width and current magnitude.

A low current, well stabilized arc can be constrained as a cylindrical shape and would look similar to geometry in Figure 5.5 below, in higher current arcs, the convection of heat forces the arc to bow upwards, an attribute that lends itself to the name ‘arcing’. The arc is made up of three main regions; the anode, the plasma column and the cathode region. The anode and cathode regions are commonly known as the electrodes where the solid metal regions transpires into a gaseous plasma.

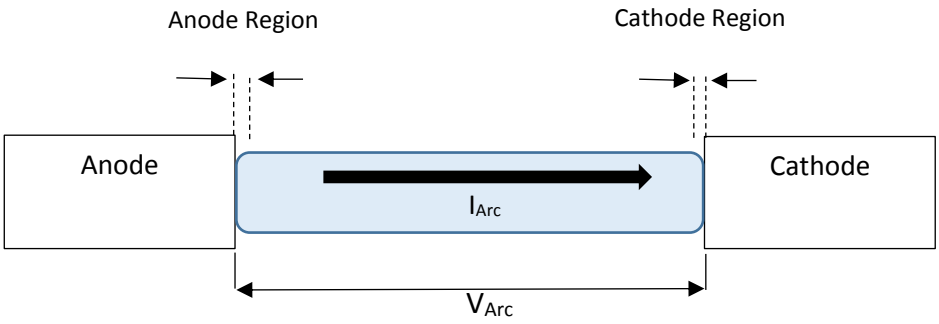


Figure 5.5. Showing the approximate geometry of a low current arc and associate fall in voltage.

Between these electrodes, a voltage drop or gradient, which is dependent upon the arc length, develops (Ammerman, 2010). Due to the complexity and nature of how the arc behaves, theoretical models become very difficult to develop that accurately describe the physics of the arcing process. This leads to most models resembling a ‘black box’ approach. Equations over the years have been developed by (Nottingham,

1926), (Ayrton, 1902), (Steinmetz, 1906), (Nottingham, 1923), (Van and Warrington, 1931), (Miller and Hildenbrand, 1973), (Hall, 1978), (Myers, Vilicheck, Stokes and Oppenlander, 1991), that approximates the fall in voltage between the anode and cathode.

5.2.2 Energy Transferred by Arcing

The energy released during arcing is subject to the law of energy conservation, and hence the electrical energy input is equal to losses encountered in the form of heat, light, pressure, sound and electromagnetic radiation. This led to the development of arc-resistance models that take into account the above to form an estimate of the electrical energy delivered during arcing (Gammon and Matthews, 2003).

For steady state dc power systems, power is defined as

$$P = V_{dc} I_{dc} \tag{123}$$

and the power for dc and single phase ac arcs can be described as

$$P_{Arc} = V_{Arc} I_{Arc} = I_{Arc}^2 R_{Arc} \tag{124}$$

as energy is a function of time, the energy associated with arcing can be approximated by

$$E_{Arc} \approx I_{Arc}^2 R_{Arc} t_{Arc} \tag{125}$$

Where time is measured in seconds.

An arc model therefore needs to incorporate the fall in voltage between the anode and cathode of the contacts, and from this, along with the power, an estimate of the energy transferred can be approximated.

(Swingler and McBride, 1996) provided a model to estimate the amount of energy brought by the electrical arc to the electrode surface under study for each time t of the breaking process under DC conditions. The input parameters such as current intensity, circuit voltage and opening velocity related to experimental conditions are used. From this, the power flux density through the arc is used to produce the output which will be used as the input to the thermal model used later on.

The anode region, the plasma region and the cathode region are modelled in terms of the arc energy transport regions.

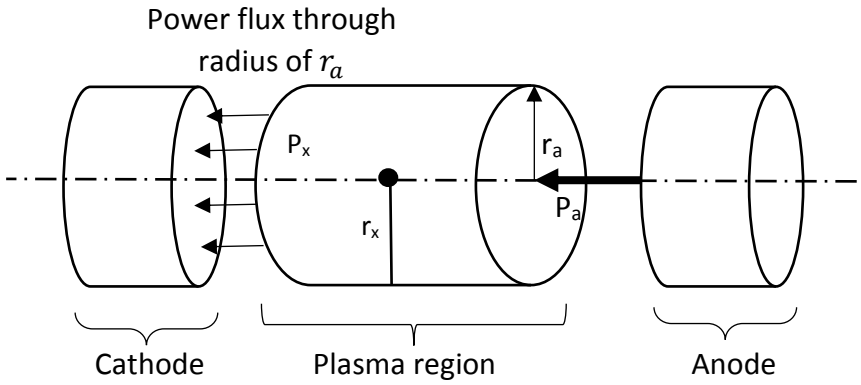


Figure 5.6. Illustrating the anode, plasma and cathode transport regions

Within each of these three regions, the power dissipation is computed from the current through the region and the voltage drop across it.

The power dissipation from the anode, through the arc region to the cathode surface is calculated by considering the two ways of energy transport processes that encompass all the mechanisms involved at a microscopic scale in the energy transport process: radial and channel transport processes. The Radial transport processes represent the processes which radiate energy equally in all directions such as thermal energy from random bombardment of particles, radiation from de-excitation of particles, etc. The channel transport processes account for mechanisms which transport energy (channel energy) toward the cathode or anode. The energy is channelled to neighbouring regions by, for instance, positive ion or electron bombardment as they are accelerated through the electric field.

This leads to the amount of energy being available from any arc region at a given time t being equal to the energy transported from any neighbouring regions by the transport processes plus the energy generated within that region itself. The power flux density out of the plasma region transporting energy towards the cathode is then given by (Swingler and McBride, 1996):

$$q_k = \frac{(P_{plasma}+P_a)}{\pi r^2} \times \left(k_1 \frac{r}{2\sqrt{r^2+r_x^2}} + k_2 \right) \qquad (126)$$

Where:

q_k is the power flux density out of any particular region transporting energy towards the cathode

K_1 is the proportion of energy radially dissipated

K_2 is the channel transport process and is assumed to be 50%

P_a is the power input at the anode region

P_{plasma} is the power dissipated in the plasma region

r_a is the radius of the arc to the axis

r_x is the distance between the point source and the plasma boundaries

Instead of K_2, another control term may be inserted, K_3, which allows the power flux density out of the plasma region transporting energy towards the anode to be found.

5.3 Heat flow, Contact Temperature and Material Erosion

In order to model the heat flow in the contact, a model of the contact is developed using the heat equation with the power input Q(r,t) given by equation (126) above.

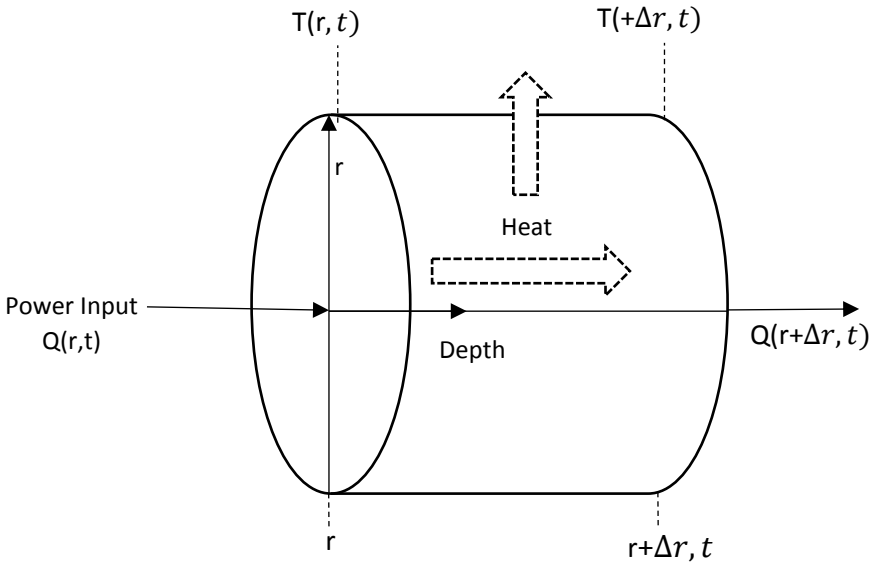


Figure 5.7. Shows the direction of heat flow Q and temperature T through the contact.

The heat equation is used to model the flow of heat from a hot to a cold area by the process of conduction, where the temperature satisfies the following equation (James, 2011).

$$\frac{1}{\alpha}\frac{\partial T}{\partial t} = \nabla^2 u \qquad (127)$$

The Fourier law for the conduction of heat states that the heat transferred across unit area is proportional to the temperature gradient. The equation (126) may be written in terms of a cylindrical coordinate system.

$$\frac{1}{\alpha}\frac{\partial T}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 T}{\partial \emptyset^2} + \frac{\partial^2 T}{\partial z^2} + \frac{\dot{q}_a}{k} \quad (128)$$

The heat flow in a cylindrical shape is only in the radial direction, hence the temperature distribution through the contact becomes $T = T(r,t)$ and the above equation (128) reduces to

$$\frac{1}{\alpha}\frac{\partial T}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) + \frac{\dot{q}_a}{k} \quad (129)$$

Where $\alpha = \frac{k}{\rho \times c}$, k is the conductivity, ρ is the density and c is the specific heat capacity of the material respectively and is often referred to as the thermal diffusivity.

5.3.1 Numerical Solution

For most problems, unless they are simple, the heat equations solution needs to revert to a numerical solution. For modelling the heat flow through the contact a numerical solution will be adopted in the form of a finite difference solution.

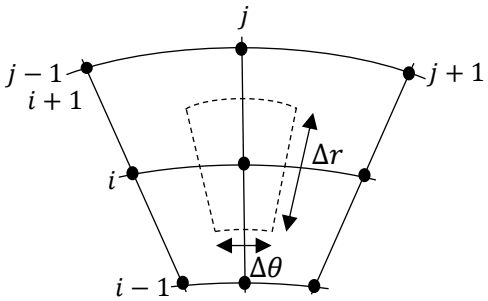


Figure 5.8. Showing the node for the finite difference solution

The radius can be determined by the index i, where $r_i = (i-1)\Delta r$, i=1,2…N and angle θ is determined by the index j, where $\theta_j = (j-1)\Delta\theta$, j=1,2…M. For time varying problems m is used to denote time, t=mΔdt where m = 0,1,… for a unit length in the z direction, the area of control volume normal to the radial direction is $\left(r - \frac{\Delta r}{2}\right)\Delta\theta$ at the inner surface and $\left(r + \frac{\Delta r}{2}\right)\Delta\theta$ at the outer surface. The area normal to the circumferential direction is Δr.

The distance between the nodes is

$$(i,j)\ to\ (i \pm 1, j): \Delta r$$

$$(i,j)\ to\ (i, j \pm 1): \Delta r$$

The difference equation can be derived from the heat equation above by considering the energy balance in the heat equation.

The heat conducted into the face between (i,j) and (i,j-1) is given by

$$= k\Delta r \frac{T_{i,j+1,m} - T_{i,j,m}}{r\Delta\theta}$$

The heat conducted into the face between (i,j) and (i,j+1) is given by

$$= k\Delta r \frac{T_{i,j+1,m} - T_{i,j,m}}{r\Delta\theta}$$

The heat conducted into the face between (i,j) and (i-1,j) is given by

$$= k(r - \frac{\Delta r}{2})\Delta\theta \frac{T_{i-1,j,m} - T_{i,j,m}}{\Delta r}$$

The heat conducted into the face between (i,j) and (i+1,j) is given by

$$= k(r + \frac{\Delta r}{2})\Delta\theta \frac{T_{i+1,j,m} - T_{i,j,m}}{\Delta r}$$

The rate at which energy is stored in the control volume is

$$= \rho cr\Delta r\Delta\theta \frac{T_{i,j,m+1} - T_{i,j,m}}{\Delta t}$$

If the heat generation is non-zero, $(\dot{q}_G \neq 0)$, then the rate of heat generation inside the control volume is $\dot{q}_G, i, j, m\ r\Delta\theta\Delta r$.

The resulting equation is

$$\rho cr\Delta r\Delta\theta \frac{T_{i,j,m+1} - T_{i,j,m}}{\Delta t} = k\left\{\frac{\Delta r}{r\Delta\theta}\left(T_{i,j+1,m} - 2T_{i,j,m} + T_{i,j-1,m}\right) + \frac{r\Delta\theta}{\Delta r}\left(T_{i+1,j,m} - 2T_{i,jm} + T_{i-1,j,m}\right) + \frac{\Delta\theta}{2}\left(T_{i+1,j,m} - T_{i-1,j,m}\right)\right\} + \dot{q}_{G,i,j,m}\, r\Delta\theta\Delta r \quad (130)$$

The equation was solved using Matlab subject to the following boundary conditions.

5.3.2 Boundary Conditions

In specifying the boundary conditions using Fourier's Law for heat conduction.

$$\ddot{q}_1 + \dot{q}_{G,1}\Delta x = -k\frac{T_{2-}T_1}{\Delta z}$$

The above equation says rate of heat conduction into the control volume combined with the rate of heat generation inside the control volume is equal to the rate of conduction out of the control volume combined with the rate storage inside the control volume.

The heat flow through the contact is subject to boundary conditions depending upon known heat flux (T∞) or convection (h) conditions. For the simulation it is assumed that some convection from the wall of the contact is present.

The initial and boundary conditions are defined as:

$$k\frac{\partial T(r)}{\partial r} + h_b T(r) = h_b T_{\infty,b} = known\ at\ r = b$$

$$\begin{cases} -k\dfrac{\partial T}{\partial z} = q_a(r,t), & z = 0 \\ \dfrac{\partial T}{\partial z} = 0, & r = R \\ T = T_0, & Z = Z_0 \\ T(r,z,0) = T_0, & t = 0 \end{cases}$$

5.4 Results from Model

Figure 5.9 below, shows the results of the heat flow through the contact for an arc duration of 10ms and 5ms, with an arc temperature of 5368.15 K. The time was predicted from the contact model in section 5.1 and the arc power was calculated using equation (124) above.

As well as the heat flow through the contact, the difference in the heat at the centre and surface of the contact due to convection is shown, the voltage distribution across the contact is also depicted using equation (12), the theoretical loss of material due to vaporization and the mass loss as a function of time.

Where the solution of the PDE is subject to fast changes, such as the temperature rise from a short arcing duration, the mesh size must be closely spaced in order to reflect the changes. To apply the finite difference method, the spatial domain is divided into M radial nodes $[0, r\,]$ each of length $\Delta r = \frac{r}{M}$, and the time domain $[0, T]$ is divided into N segments, each of duration $\Delta t = \frac{T}{N}$. To ensure stability(s) in the

solution $S = \alpha \frac{\Delta t}{\Delta r^2} \leq \frac{1}{2}$, where α is the constant relating conductivity, specific heat capacity and thermal diffusivity respectively. This implies that as the radial spatial interval Δr is decreased for increased accuracy, the time step Δt must be decreased at the cost of more computations or else there is a risk of instability. The mesh size was made equal to the radius of the contact divided by the radius nodes, and the time step selected to satisfy the condition *S*.

The results from the model are shown in figures 5.9 – 5.18, respectively. Figures 5.9 and 5.14 shows the cylindrical shaped contact with section taken parallel to its axis. The heat distribution of the initial arc temperature on the face of the contact is plotted in the z direction with respect to time. As may be expected, the temperature is greater as the time is increased, representing more power being coupled into the contact through the energy from the arc.

As convection losses were also modelled, the temperature at the centre of the contact and surface is also shown in the graphs of 5.10 and 5.15 where there is a substantial differential at the initial time when the arc is applied. This temperature difference narrows quickly as the heat starts to distribute through the contact, hence there is a higher localised temperature towards the centre of the contact and greater material loss is liable to take place within this region.

The voltage temperature relationship given by the Wiedemann-Franz law is shown in figures 5.11 and 5.16, showing the relationship between the temperature and the voltage across the contact. The graphs concur the higher the temperature the more voltage is present.

The last two sets of figures show the estimations of mass losses due to the vaporisation process due to enthalpy from self-diffusion. In figures 5.12 and 5.17, the estimated segment of mass loss against time is shown. In each case the initial estimate is based upon the mass of the contact from the dimension and density of the material. The process is iterative, with the mass loss due to vaporisation being calculated for each step, along with temperature at that step, and an update of the contact mass and thermal properties. This was not available in commercial software at the time of writing and was the reason for writing a program to model the heat flow in the contact. The results show the loss of material over time in figures 5.13 and 5.18, the value equates well with the estimations of (Rieder and Weichsler, 1992) discussed in section 5.6.

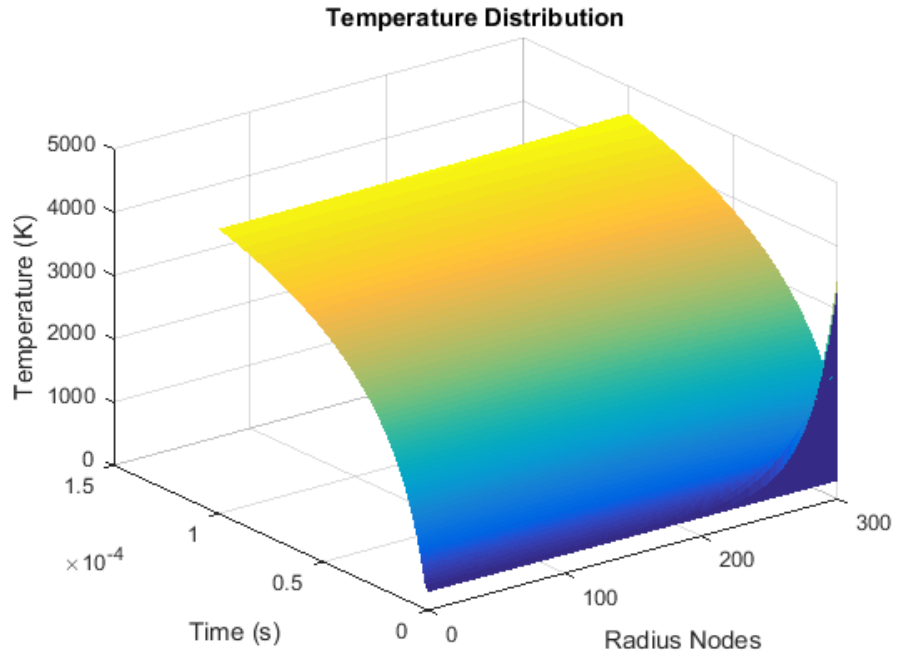


Figure 5.9. Showing the heat distribution through the contact at a time of 10ms.

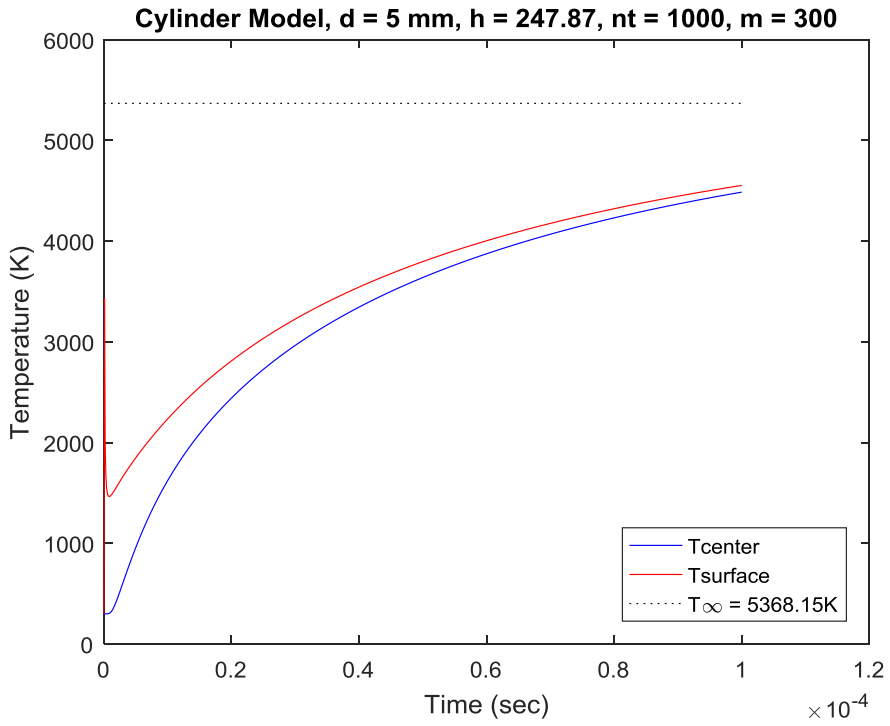


Figure 5.10. Showing the temperature at the center and surface differential due to convection at 10ms.

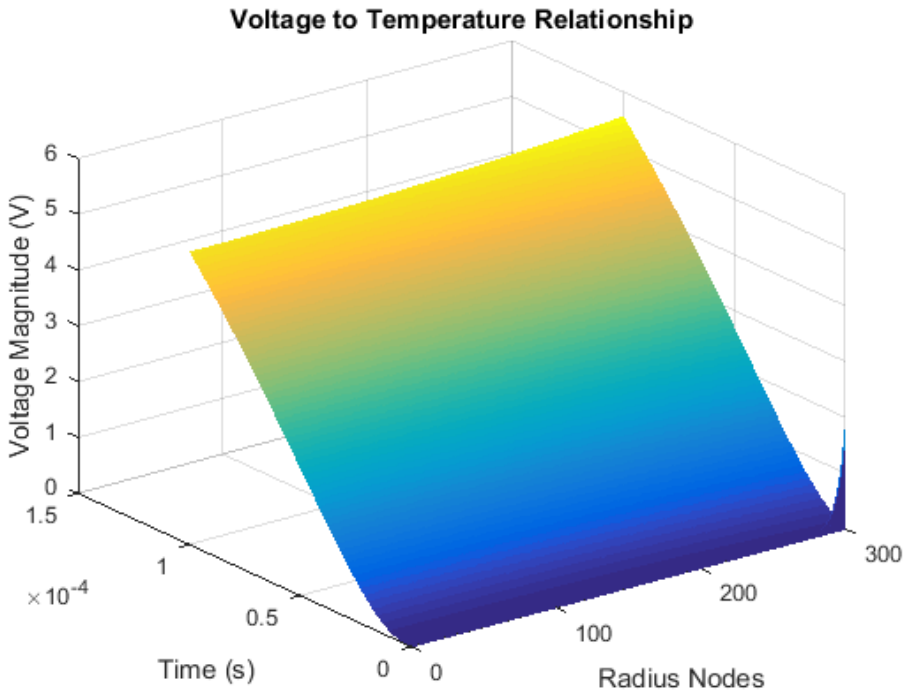


Figure 5.11. Illustrating the voltage-temperature relationship corresponding to the Wiedemann–Franz Law.

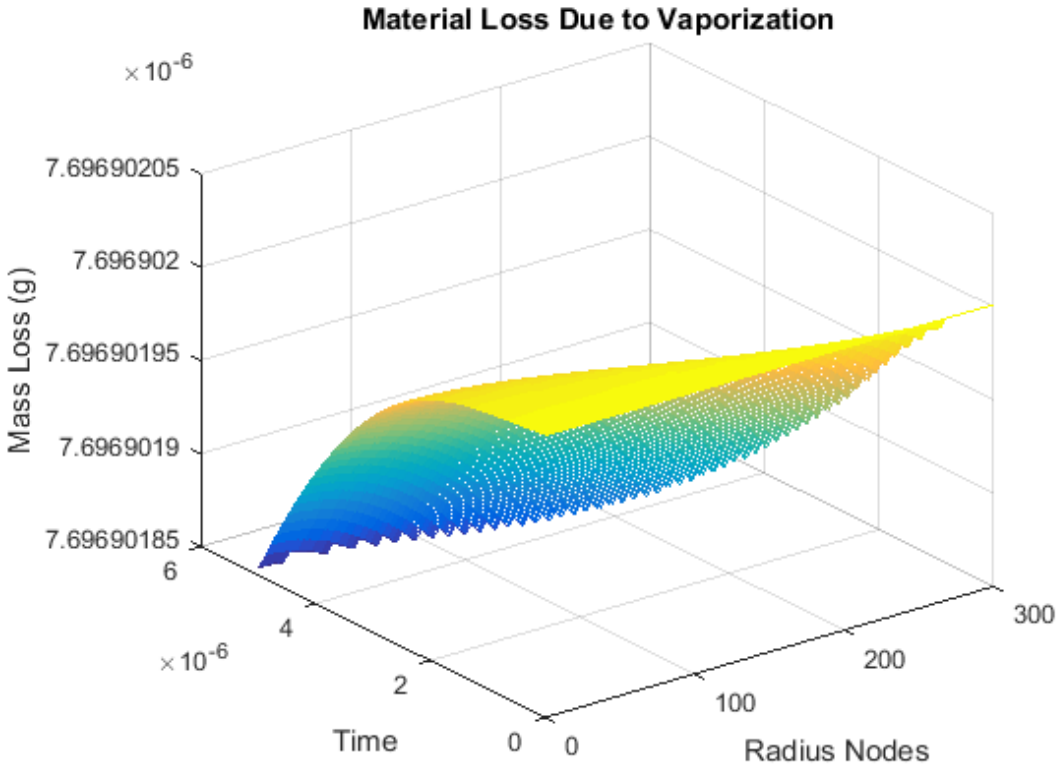


Figure 5.12 Material loss visualization due to vaporization of the contact surface.

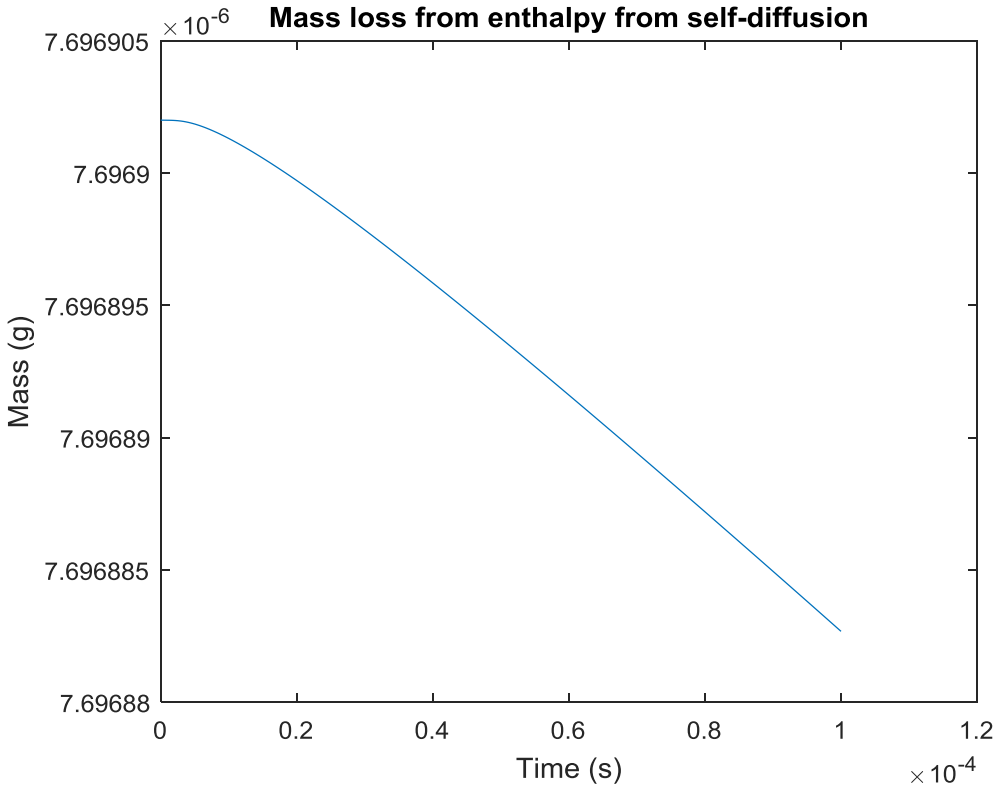


Figure 5.13. Material loss from contact for 10 ms Arc.

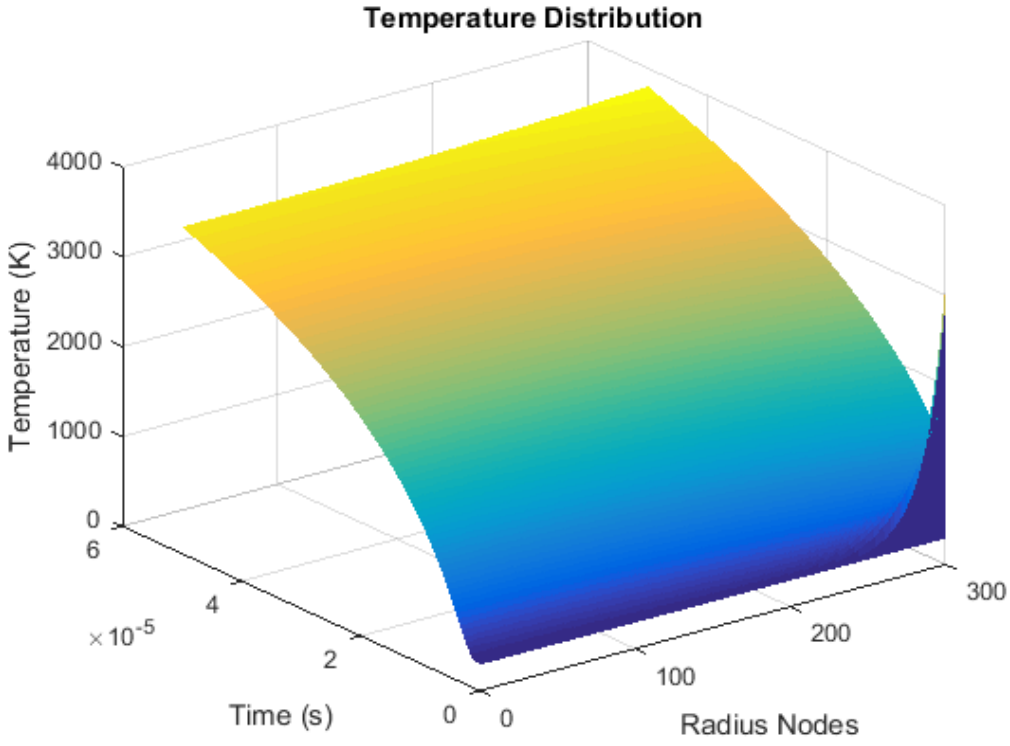


Figure 5.14. Showing the heat distribution through the contact at a time of 5ms.

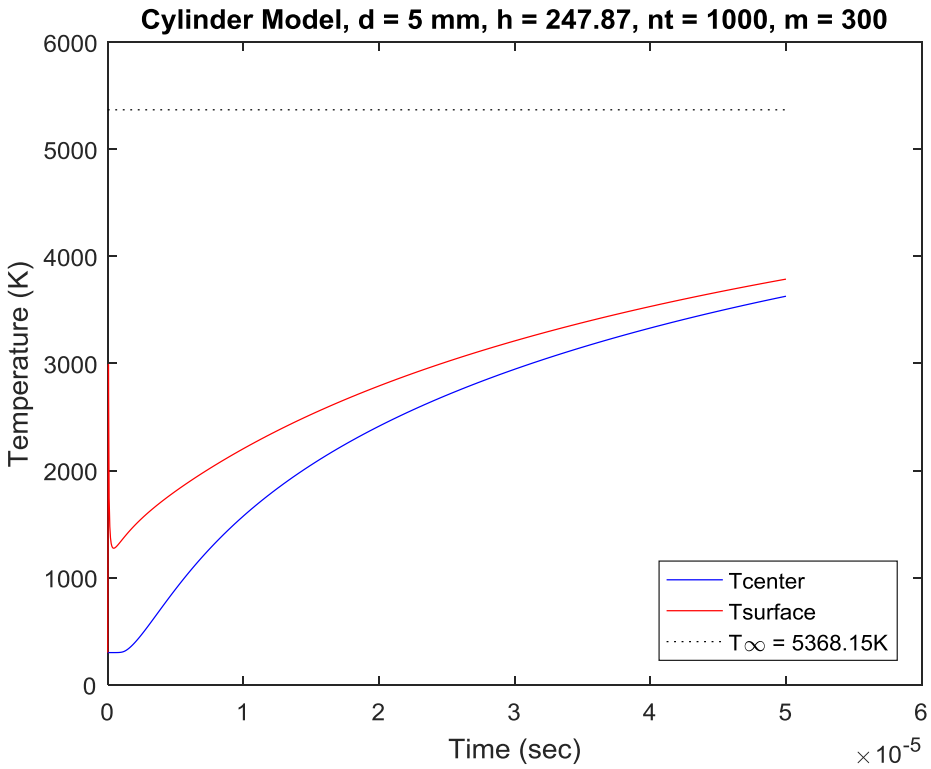


Figure 5.15. Showing the temperature at the center and surface differential due to convection at 5ms.

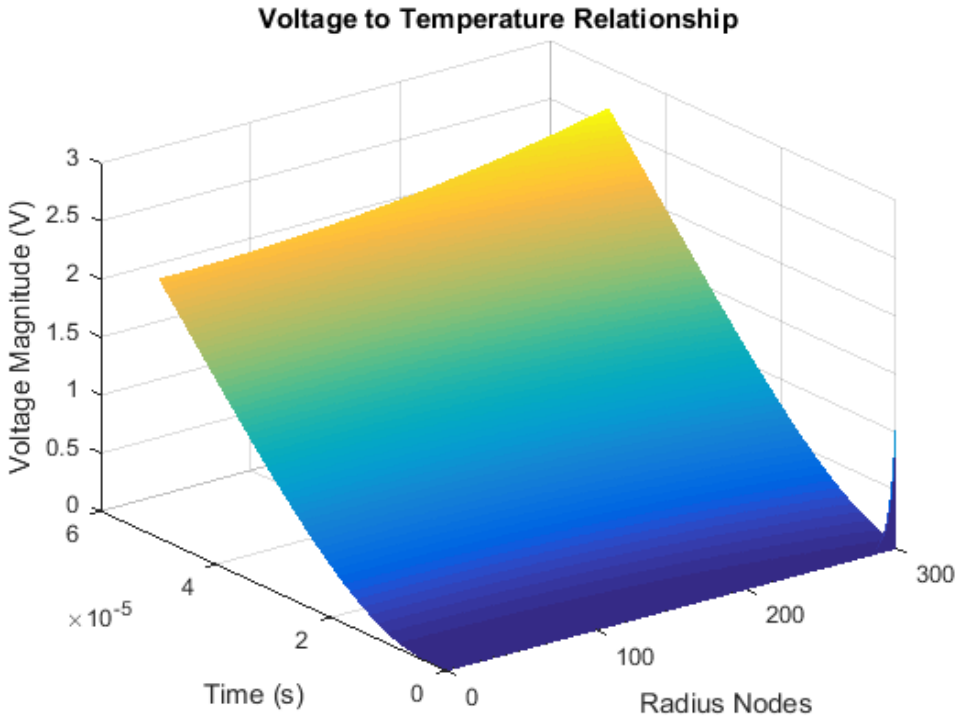


Figure 5.16. Illustrating the voltage-temperature relationship corresponding to the Wiedemann–Franz Law.

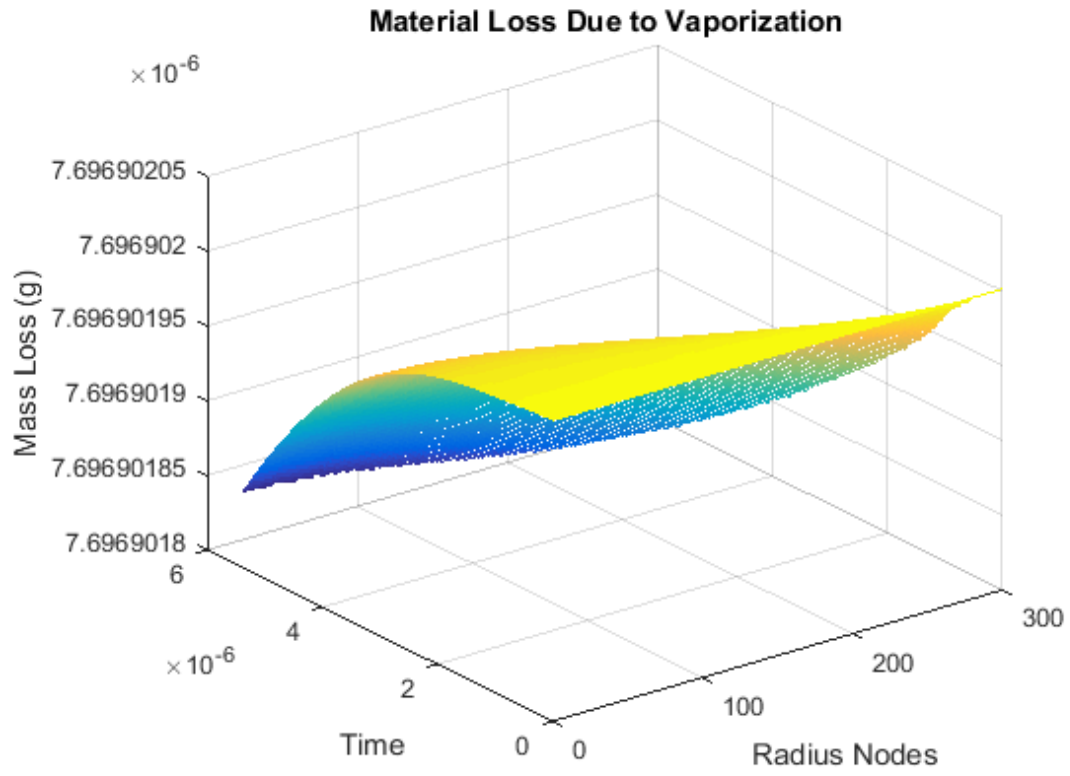


Figure 5.17. Material loss visualization due to vaporization of the contact surface.

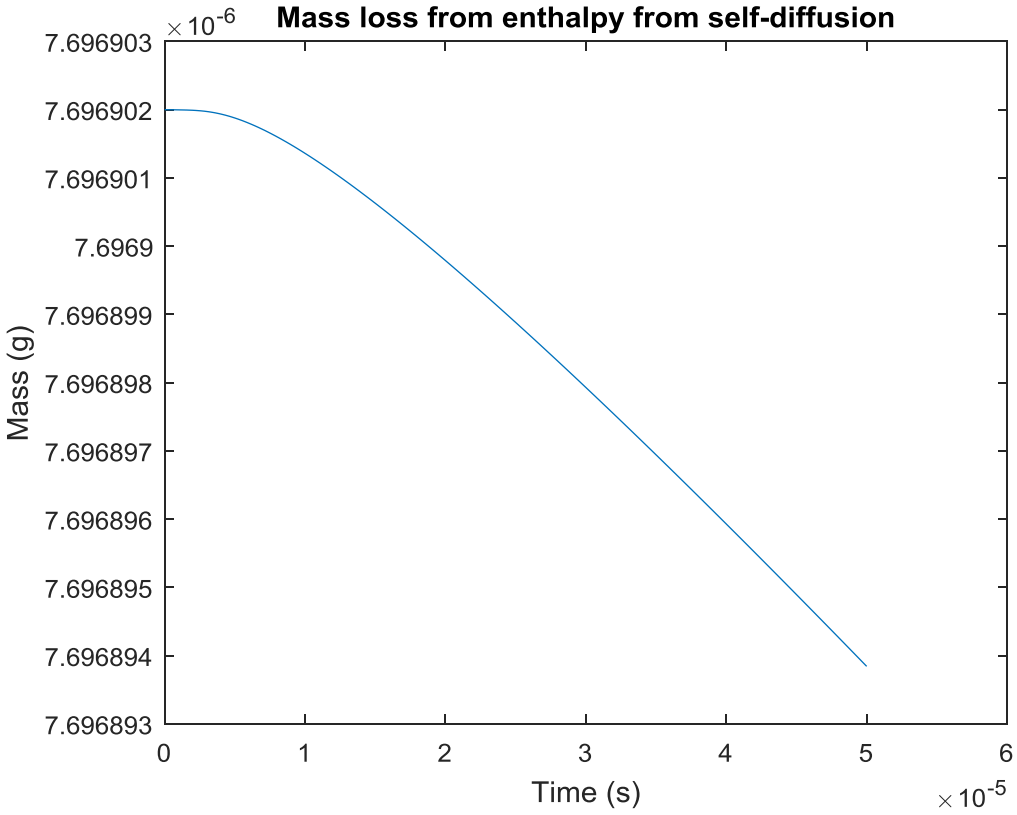


Figure 5.18. Material loss from contact for 5 ms Arc.

5.5 Mass Loss

One parameter that affects the rate of erosion is the conduction of heat away from the contact surface. The dimensions of the contact e.g. area, length, breadth and thickness all play an important role in the rate of erosion assessment.

Mass loss estimation from the contact surface has tended to be by experimental observation (Mcbride, 2000), (Swingler, 2010), a lot of work was done in the pre digital 1950s and 60s when telephone exchanges used relays.

Tables are available giving constant values (Holm, 2000), for mean values of coefficients characterizing the arc material transfer on making or breaking contact during a long series of operations, ranging from 0.03 to 1.1. For the relay being used in this work, the contact material is a silver alloy consisting of Silver (Ag) and tin oxide (SnO_2). Hence from the table a material transfer rate γ_p = 0.6 and the loss due to evaporation from arcing is β = 0.8.

Other assessments of mass loss come from (Rieder and Weichsler, 1992), who plotted the make and break Mass loss/arcing time, mg.s–1 in AC contacts against number of operations. (Turner & Turner, 1966 & 1968) extrapolated their results for various contact materials to come up with the following equation

$\frac{d\omega}{dt} = k_1 I^{1.6}$ μg.s^{-1}

Where DEC is the Discontinuous Erosion Current, the parameter k_1 may be estimated from:-

Contact	K (I<DEC)	K (I>DEC)
Ag	0.8	25
Cu	2.4	36
AgCdO	0.4	6

Where the discontinuous erosion current is either above 1000 A or below 1000 A.

Mass loss in this model is due to vaporization of the contact surface from the arc and hence the mass loss is assumed to be due to diffused metal. The coefficient for any given metal shows the Arrhenius temperature dependence and may be written as a function of temperature according to the equation (Alcock, 2000).

$$D = D_0 \exp(-\frac{\Delta H^*}{RT}) \quad (131)$$

Where D_0 is a constant for the contact material, and $\Delta H^*$is called, by analogy with the Arrhenius equation for gas reaction kinetics, the ‘activation energy’, in this case it is the vaporization energy for the contact material, R is the universal gas constant and T is the temperature respectively. The pre-exponential term has a fairly constant range of values for most metals of approximately 0.1 to 1, and the activation energies follow the same trend as the heats of vaporization (Alcock, 2000). The vaporization causes vacancy migration due to the diffusion and the activation enthalpy for self-diffusion $\Delta H^*$is made up from sum of the energy to form a vacancy ΔH_{Vac} and the energy to move the vacancy ΔH_{Diff}.

The initial mass is calculated for the dimensions of the contact and this is updated at each iteration, via recalculation of the mass due to loss, as well as the effective heat capacity and thermal conductivity. Therefore the thermal effects and mass loss are representative as time progresses.

5.5.1 Model of Contact Degradation Based on a Feature Vector Failure Parameter

In order to represent the damage as it evolves, a mathematical model is needed. The process is deemed to be Markovian in nature, each damage state evolving from the previous damage incurred. In this case, one approach is to take a measurable feature that represents the damage occurring, e.g. the contact resistance, and propagate this in terms of how the 'feature of the damage grows'. Figure 5.19 below shows the methodology in the prediction of the RUL. The accelerated aging experiment is used to degrade a parameter, in this case the contact resistance. The parameters degradation is modelled, this model is formed from the work done from the above to extract the mass loss and will be related to the contact resistance by using the equation $R_c = \frac{\rho}{2a}$ this describes the relation between the apparent surface area of the contact and the resistivity of the material. Finally the model will be used to produce an estimation of the RUL of the contact and compared to the actual degradation data extracted experimentally. The above process is summerised in the diagram below.

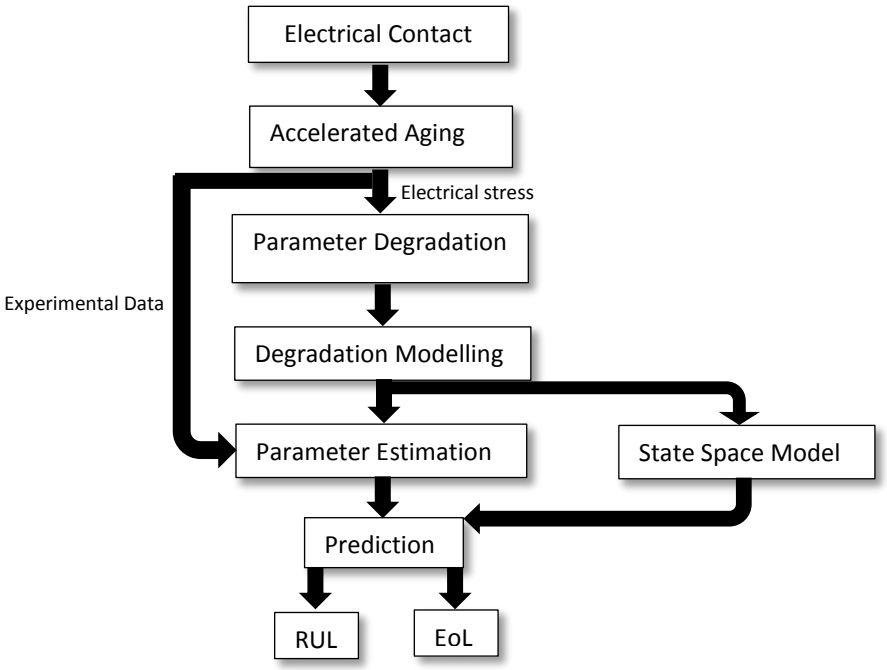


Figure 5.19. The processes of forming a PoF model for estimation of RUL.

5.6 Physics Based State Space Model

In determining a physics based model and due to the multi-faceted failure mode, the assumed dominant parameter, in this case arcing has been used. The model is centered on arc erosion and mass loss from the contact volume. This mass loss will affect the effective surface area of the contact, decreasing the apparent area between the two contacts and hence increasing the contact resistance.

The volume of the contact is equal to

$$V_c = \pi r^2 h_c \tag{132}$$

where r is the contact radius and h_c is the height of the contact.

At any time the change in volume of the contact may be related to mass loss of the contact where the mass loss is due to $A_s \times \omega \times V_m \times t$

Where: A_s is the Holm surface area (m^2)

ω is the rate of mass loss which is equal to $\frac{mass\ loss}{surface\ area \times time}$ (Kg$m^{-2}s^{-1}$)

V_m is the volume of contact material molecule (m^3/mol)

t is the time of the arc duration (s)

Hence,

$$V_c(t) = A_s \times \omega \times V_m \times t \tag{133}$$

The effective area of the surface can be represented to good approximation by the Holm radius (i.e., the Holm circle) of the area over which electrical contact occurs

(Holm, 2000). This results suggest that the details of the number and spatial distribution of the *a*-spots are not important to the evaluation of contact resistance in many practical applications where electrical contact occurs reasonably uniformly over the nominal contact area, that is, in the absence of electrically insulating surface films. As a first approximation, the Holm radius may be estimated from the true area of contact, *A*, as $\left[\frac{A}{\pi}\right]^{1/2}$ (Slade, 2013).

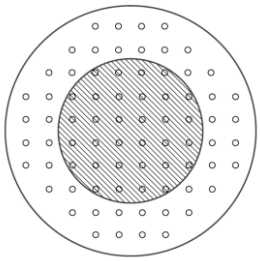


Figure 5.20. Regular array of *a*-spots; the shaded area is the single continuous contact with the same resistance; the outer circle is the Holm radius of the cluster.

From Chapter 2, the contact resistance can be approximated by $R_c = \frac{\rho}{2a}$ where ρ is the effective resistivity (Ωm) and a is the diameter of the a-spot.

Replacing a in $R_c = \frac{\rho}{2a}$ above gives the contact resistance in terms of the effective area.

$$R_c = \frac{\rho}{2\left[\frac{A_s}{\pi}\right]^{1/2}} \tag{134}$$

From equation (131) above A_s may be replaced

$$R_c(t) = \frac{\rho}{2\left[\frac{\frac{V_c(t)}{\omega \times V_m \times t}}{\pi}\right]^{1/2}} = \frac{\rho}{2\left[\frac{V_c(t)\pi}{\omega \times V_m \times t}\right]^{1/2}}$$

5.6.1 Dynamic Model Development

The volume of the contact may also be represented by $V_c = A_s \times h_m$, hence the contact surface area can be represented by the volume of the contact and the height of the contact material removed h_m.

$$A_s = \frac{V_c}{h_m} \tag{135}$$

Therefore,

$$R_K = \frac{\rho}{\frac{2V_c}{h_m}} = \frac{\rho h_m}{2V_c} \tag{136}$$

The first order discrete approximation for change in contact volume can be given by

$$\frac{dV_c}{dt} = A_s \times \omega \times V_m$$

$$V_{c(k+1)} = V_{c(k)} + \frac{dV_c}{dt}\Delta t$$

$$V_{c(k+1)} = V_{c(k)} + A_s \times \omega \times V_m \times \Delta t \tag{137}$$

From equation (137) we have

$$V_{c(k)} = \frac{\rho h_m}{2R_{c(k)}}$$

Putting $\alpha = \frac{\rho h_m}{2}$ gives

$$V_{c(k)} = \frac{\alpha}{R_{c(k)}}$$

Hence, equation (137) may be rewritten as,

$$\frac{\alpha}{R_{c(k+1)}} = \frac{\alpha}{R_{c(k)}} + \frac{dR_{c(k)}}{dt}\Delta t$$

$$\frac{\alpha}{R_{c(k+1)}} = \frac{\alpha}{R_{c(k)}} + A_s \times \omega \times V_m \times \Delta t$$

Rearranging and inverting,

$$R_{c(k+1)} = R_{c(k)} + \frac{\alpha}{A_s \times \omega \times V_m} \times \Delta t \tag{138}$$

Replacing A_s by $\left[\frac{A_s}{\pi}\right]^{1/2}$ and α

The complete discrete time dynamic model for contact resistance degradation can be summarized as the following linear state model:

$$R_{c(k+1)} = R_{c(k)} + \frac{\frac{\rho h_m}{2}}{\left[\frac{A_s}{\pi}\right]^{1/2} \times \omega \times V_m} \times \Delta t \tag{139}$$

The results from the model are projected using a Particle filter (see section 3.6.3) along with employing prognostic metrics, discussed below to assess the effectiveness of the model with the different data sets.

5.7 Prognostics Performance Metrics

The task of prognostics is to provide an estimate of the time from the current state, to the point of failure. This prognosis also needs to take into account the operating conditions and how the system is to be used in the future in order to give a RUL prediction and user beneficial information that allows decisions to be made based on safety and cost, depending on the priority of the operational use.

This prediction is typically carried out once the component or system has deteriorated to a point where a defined failure threshold is reached. This however, does not mean that failure has occurred, but the system or component is deviating from its normal operating range or specification, and this damage often grows into a complete failure under future operating conditions. Continuous monitoring of the feature vector may be achieved by the comparison of the current measurement with a threshold defining the onset of failure within the normal band of operation as in figure 5.21.

How accurate this prediction is has started to be addressed; in the absence of defined metrics for assessing prognostic performance, new methodologies have started to be developed (Saxena, 2008). However, at present the performance evaluation methods are based around off-line performance, and hence are applicable for where run-to-failure data and true End-of-Life (EoL) is known in advance. Whereas on-line metrics are in development, they are still very much in their infancy. Therefore, off-line methods at present are used for verification and can, during the algorithm development phase, be useful to fine-tune prognostic algorithms.

5.8 Prognostic Categorisation

5.8.1 Metrics for Prognostic Applications

There are numerous metrics in which the performance of a prognostic methodology may be categorised (Saxena, 2010); namely certification metrics, cost benefit metrics, Mean time before failure MTBF to mean time before usage replacement MTBUR ratio, Life Cycle Cost (LCC), Return on Investment (ROI), Technical Value and Total Value.

As well as the above, which are mainly cost based performance metrics, performance based metrics such as computational performance and metrics for reliability analysis have also been implemented in literature (Saxena, 2008).

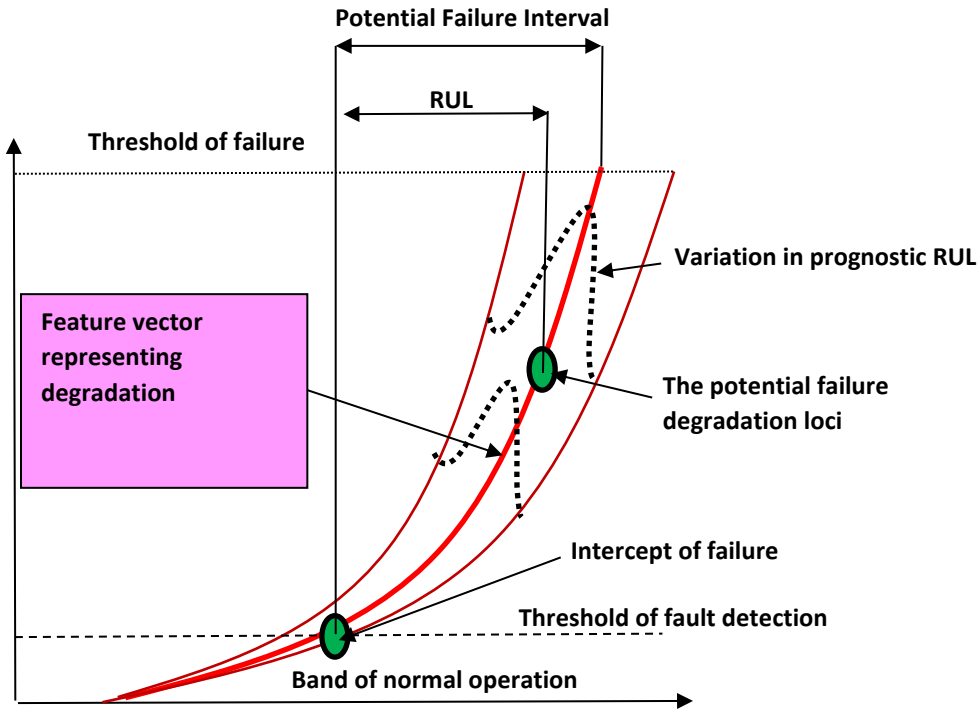


Figure 5.21. The prognostic framework

5.9 Metrics for Prognostics Algorithm Performance

Uncertainties arise from various sources in a PHM system (Coppe, 2009), (Hastings and Mcmanus, 2004) & (Orchard et al., 2008). Some of these sources include:

• Model uncertainties (errors in the representation and parameters of both the system model and fault propagation model),

• Measurement uncertainties (these arise from sensor noise, ability of sensor to detect and disambiguate between various fault modes, loss of information due to data pre-processing, approximations and simplifications),

• Operating environment uncertainties,

• Future load profile uncertainties (arising from unforeseen future and variability in usage history data),

• Input data uncertainties (estimate of initial state of the system, variability in material properties, manufacturing variability), etc.

Performance metrics for prognostics can be classified into accuracy, precision, and robustness. The definition of accuracy is the degree of closeness of the prediction to the actual failure time. Where the precision is defined as the spread of prediction performed at the same time and the robustness is the sensitivity of the predictions changes of algorithm parameter variations or external disturbances. A comprehensive list of performance metrics, is given by (Saxena, et al., 2008), (Saxena, et al., 2009) & (Saxena, et al., 2009) of which the most prevalent are discussed below.

5.9.1 Performance Metrics

Over the years, various metrics have been developed for the measure of accuracy, namely; Average Bias, Mean Square Error (MSE), Root Mean Squared Error (RMSE) and Median absolute percentage error (MdAPE).

Average bias is a conventional metric that has been used in many ways as a measure of accuracy as well as standard deviation which allows the dispersion/spread of the error with respect to the sample mean of the error to be realised. However, simple average bias metric suffers from the fact that negative and positive errors cancel each other and high variance may not be reflected in the metric. Therefore, MSE averages the squared prediction error for all predictions and encapsulates both accuracy and precision. A derivative of MSE, often used, is Root Mean Squared Error (RMSE).

MAPE weighs errors with RULs and averages the absolute percentage errors in the multiple predictions. For prediction applications it is important to differentiate between errors observed far away from the EoL and those that are observed close to EoL. Smaller errors are desirable as EoL approaches.

$$MSE = \frac{1}{l}\sum_{i=1}^{l} \Delta(y_i - f_i)^2 \tag{140}$$

$$MAPE = \frac{1}{l}\sum_{i=1}^{l} \left|\frac{100\Delta(y_i - f_i)}{n y_i}\right| \tag{141}$$

Where y_i : is the actual RUL value at time point 'i'

f_i : is the predicted RUL value at time point 'i'

n :The number of prediction time points

One of the major downfalls of the above metrics is that they are not designed for applications where RULs are continuously updated as more data is available. It is desirable to have metrics that can characterize improvement in the performance of a prognostic algorithm as time approaches near end-of-life. The next section describes the work of the extensive studies carried out by the NASA IVHM research group to solve this problem.

5.9.2 Prognostic Horizon (PH)

The Prognostic Horizon is defined as the range between the points where the predictions fall under the allowable error bound (α) for the first time and the end-of-life time point. This metric basically shows that the predicted estimates are within specified limits around the actual EoL and may be considered trust worthy within these bounds. While comparing algorithms, an algorithm with longer prediction horizon would be preferred.

$$H = EoP - i \tag{142}$$

where $i = min\left\{j \middle| (j \in l) \wedge \left(r_*(1-\alpha) \leq r^l(j) \leq r_*(1+\alpha)\right)\right\}$

i : The first time index when predictions satisfy α-bounds

α : Accuracy modifier

EoL : The ground truth end-of-life

ℓ : Set of all RUL estimation point time indexes

l : Test sample or specimen number

$r*$: Actual RUL

(j) : Predicted RUL at time instance 'j' for the test sample number 'l' (i.e. can be mean or median of prediction RUL distribution).

For instance, a PH with error bound of α = 20% identifies when a given algorithm starts predicting estimates that are within 20 % of the actual EoL.

5.9.3 α-λ Accuracy

It may be of interest whether the prediction is within a specified accuracy level at a particular time, this metric provides a way to quantify the prediction at certain time instances. For example it may be required that a prediction falls within 20% accuracy (i.e., α=0.2) halfway to failure from the time the first prediction is made (i.e., λ=0.5).

λ is the time and may be specified as the percentage of total remaining life from the point the first prediction is made or a given absolute time interval before EoL is reached. For the tests carried out in this thesis, α-λ *accuracy* is defined as the prediction accuracy to be within α*100% of the actual RUL at specific time instance t_λ expressed as a fraction of time between the point when an algorithm starts predicting and the actual failure. For example,

$$[1-\alpha]r_*(t) \leq r^l(t_\lambda) \leq [1+\alpha]r_*(t) \quad (143)$$

where α :accuracy modifier

λ :time window modifier

$t_\lambda = t_p + \lambda(EOL - t_p)$

λ : Time window modifier

t_p : Prediction time

The higher the percentage the better the ability of the algorithm is at prognosis.

5.9.4 Relative Accuracy (RA)

Relative Accuracy is similar to the α-λ accuracy metric. However, instead of finding out whether the predictions fall within a given accuracy levels at a given time instant, the accuracy level is measured. The time instant is again described as a fraction of actual remaining useful life from the point when the first prediction is made. An algorithm with higher relative accuracy is desirable.

$$RA_\lambda = 1 - \frac{|r_*(t_\lambda) - r^l(t_\lambda)|}{r_*(t_\lambda)} \quad (144)$$

where $t_\lambda = t_p + \lambda(EOL - t_p)$

5.10 Physics Based Model Results

Sample 1

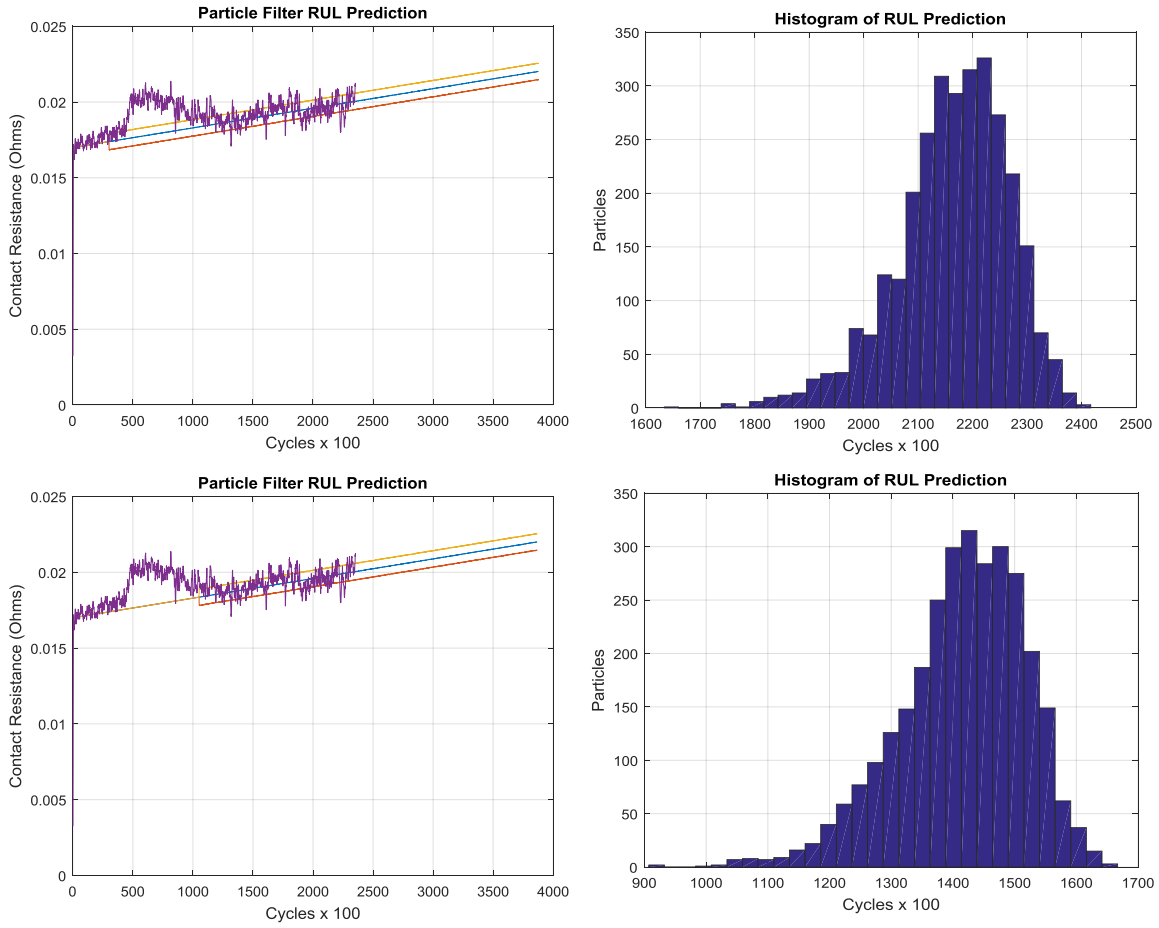
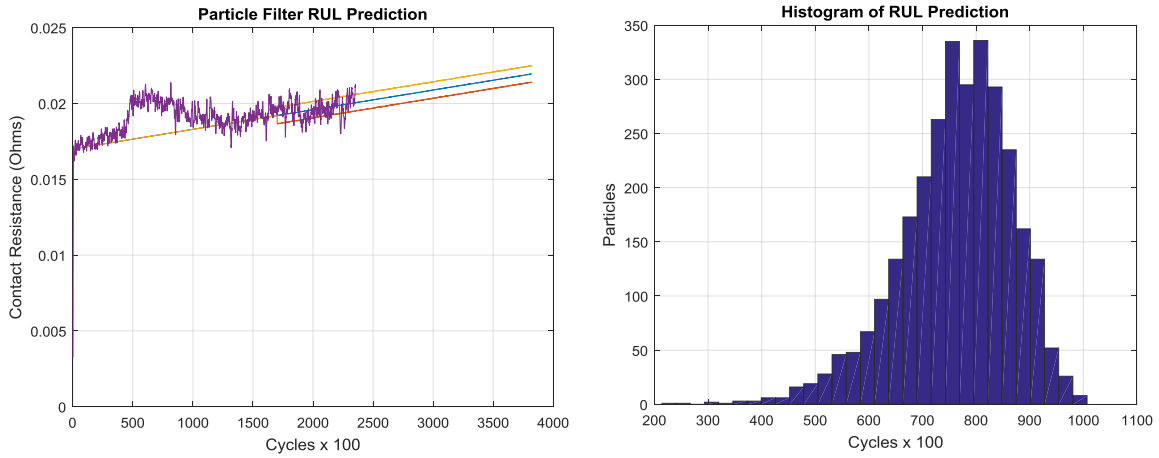


Figure 5.22. Illustrating the particle filter projection at 300, 1100 and 1700 cycles respectively and the RUL histograms in sample 1.



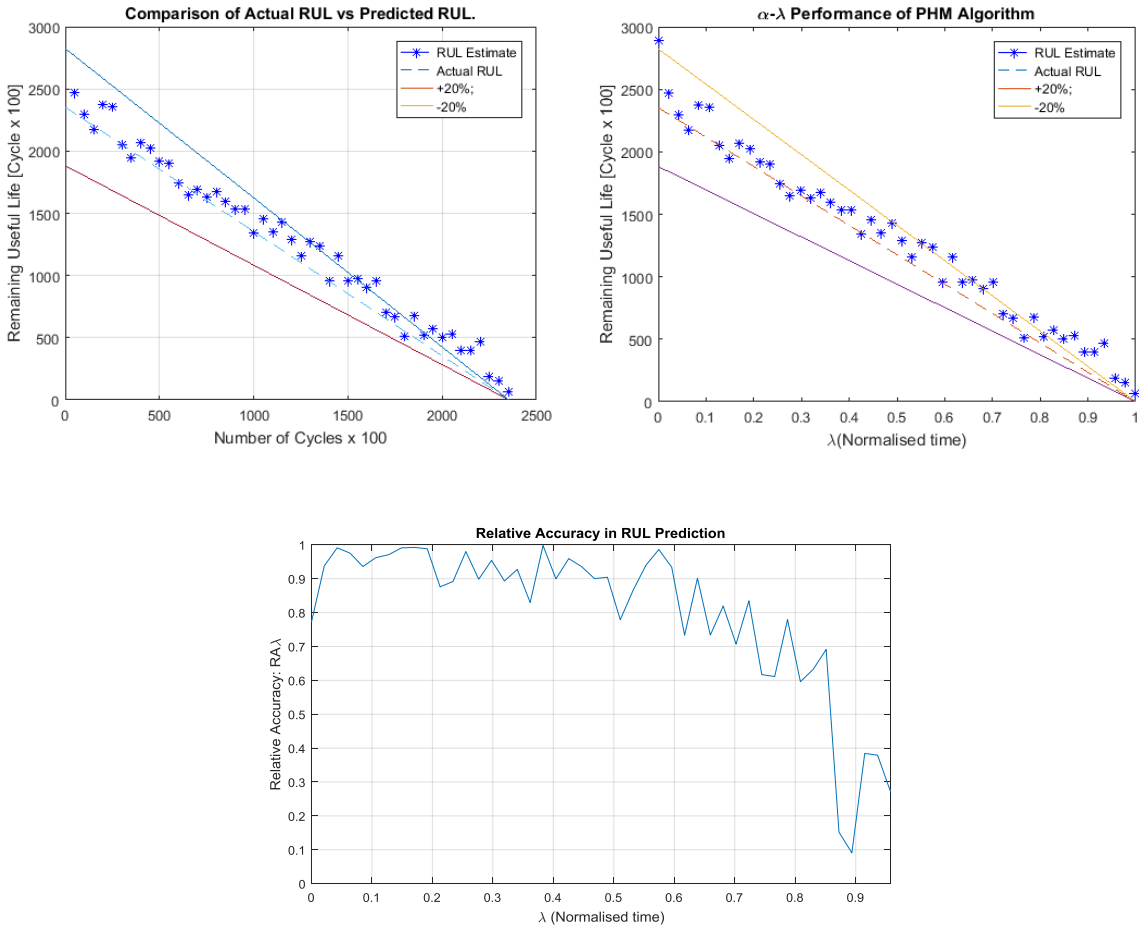
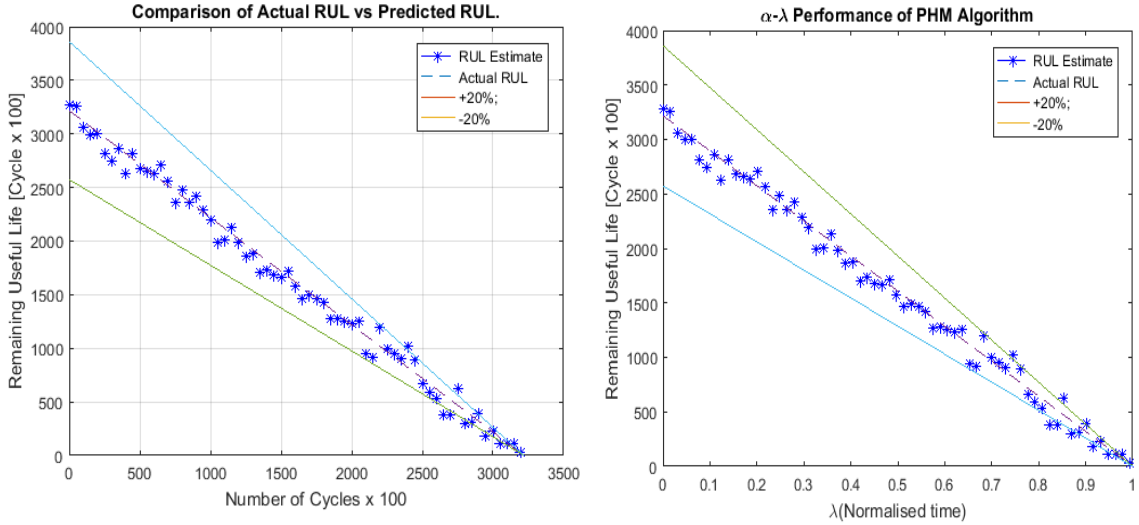


Figure 5.23. Sample 1. RUL, alpha-lambda and relative accuracy respectively, for sample 1.

Sample 2



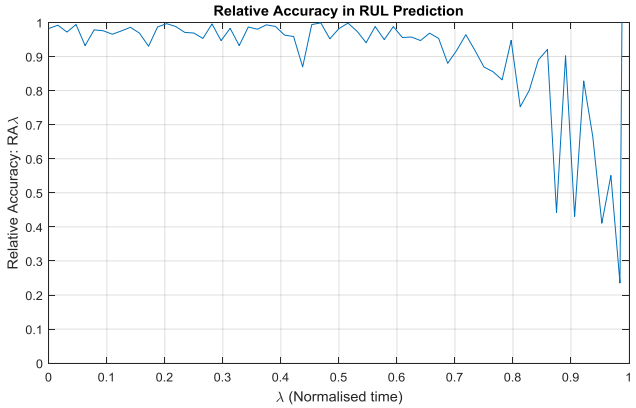


Figure 5.24. Sample 2. RUL, alpha-lambda and relative accuracy respectively, for sample 2.

Sample 3

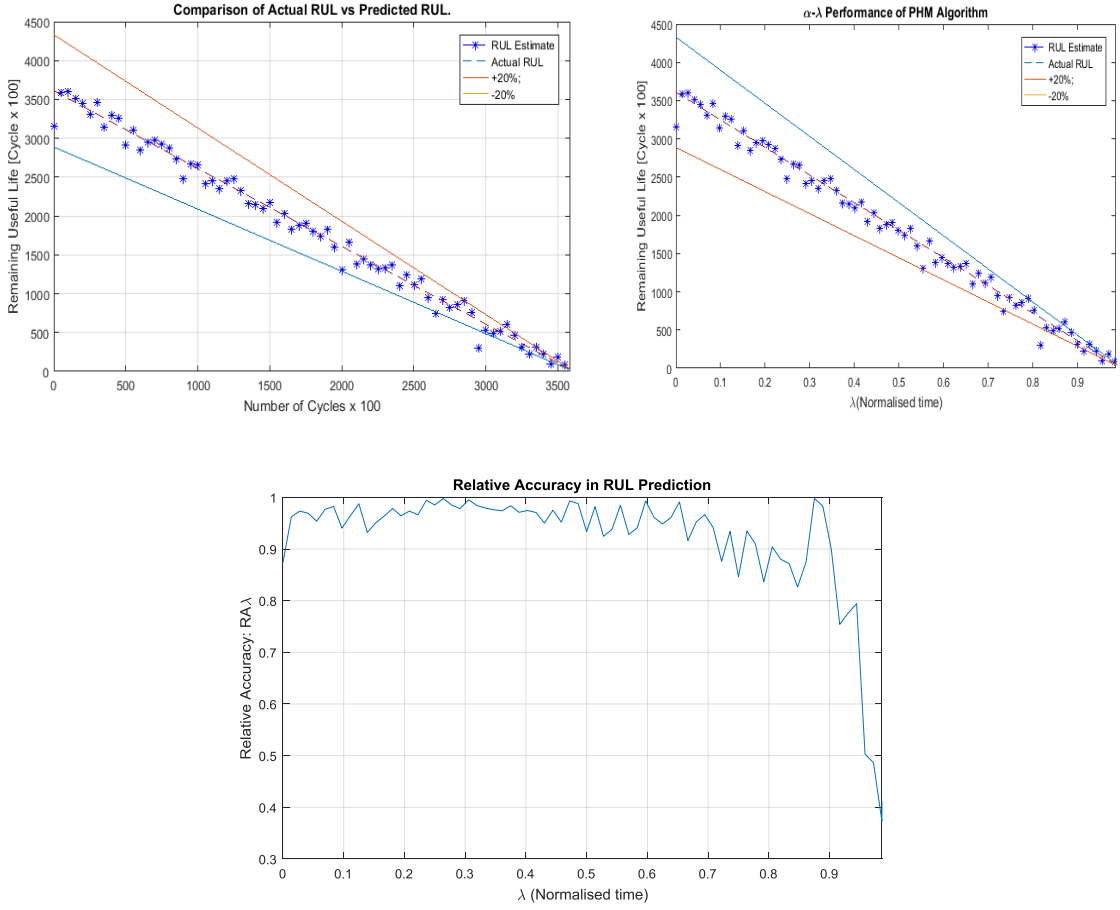


Figure 5.25. Sample 3. RUL, alpha-lambda and relative accuracy respectively, for sample 3.

Sample 4

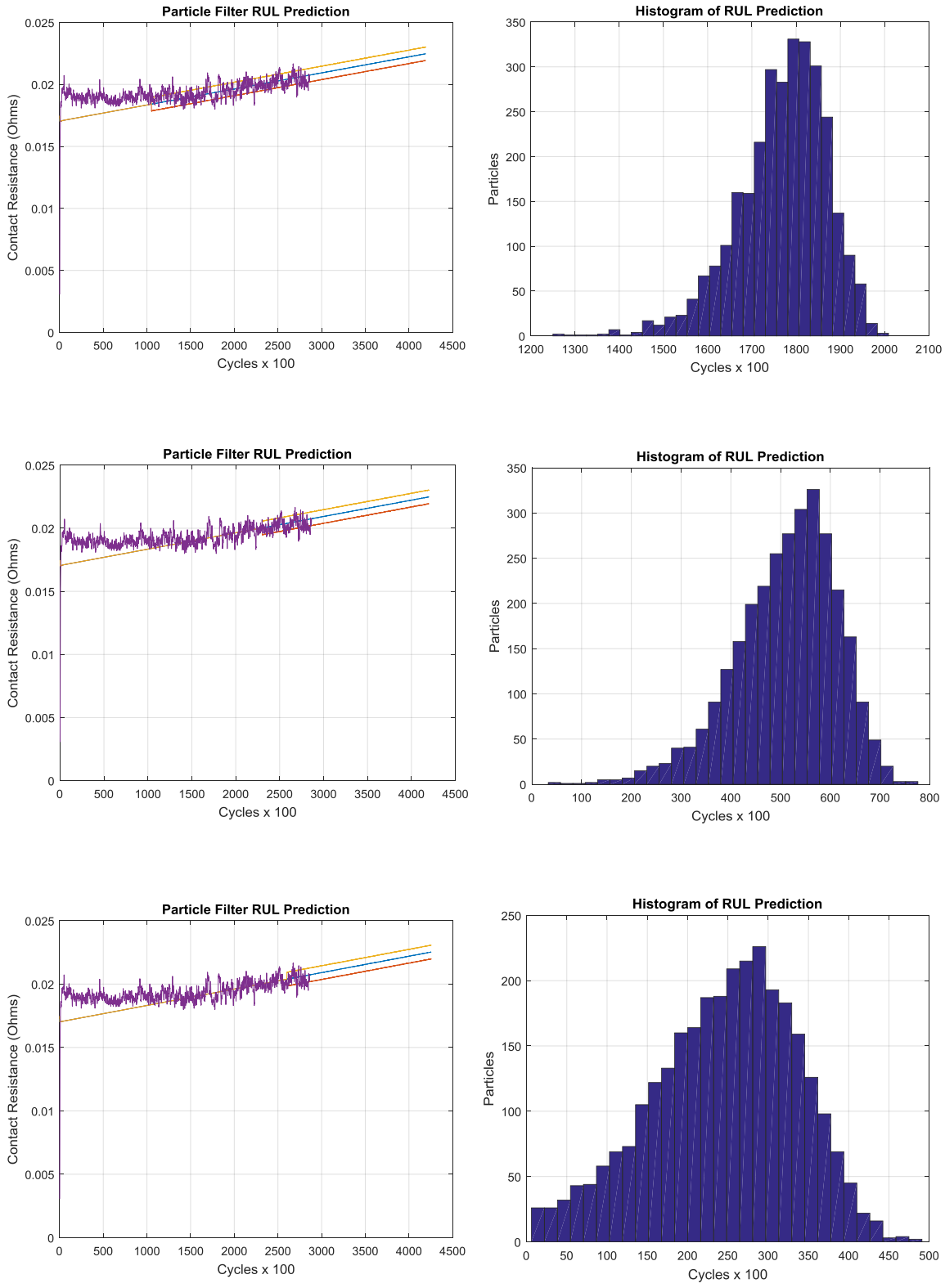


Figure 5.26. Illustrating the particle filter projection at 1100, 2200 and 2700 cycles respectively and the RUL histograms in sample 4.

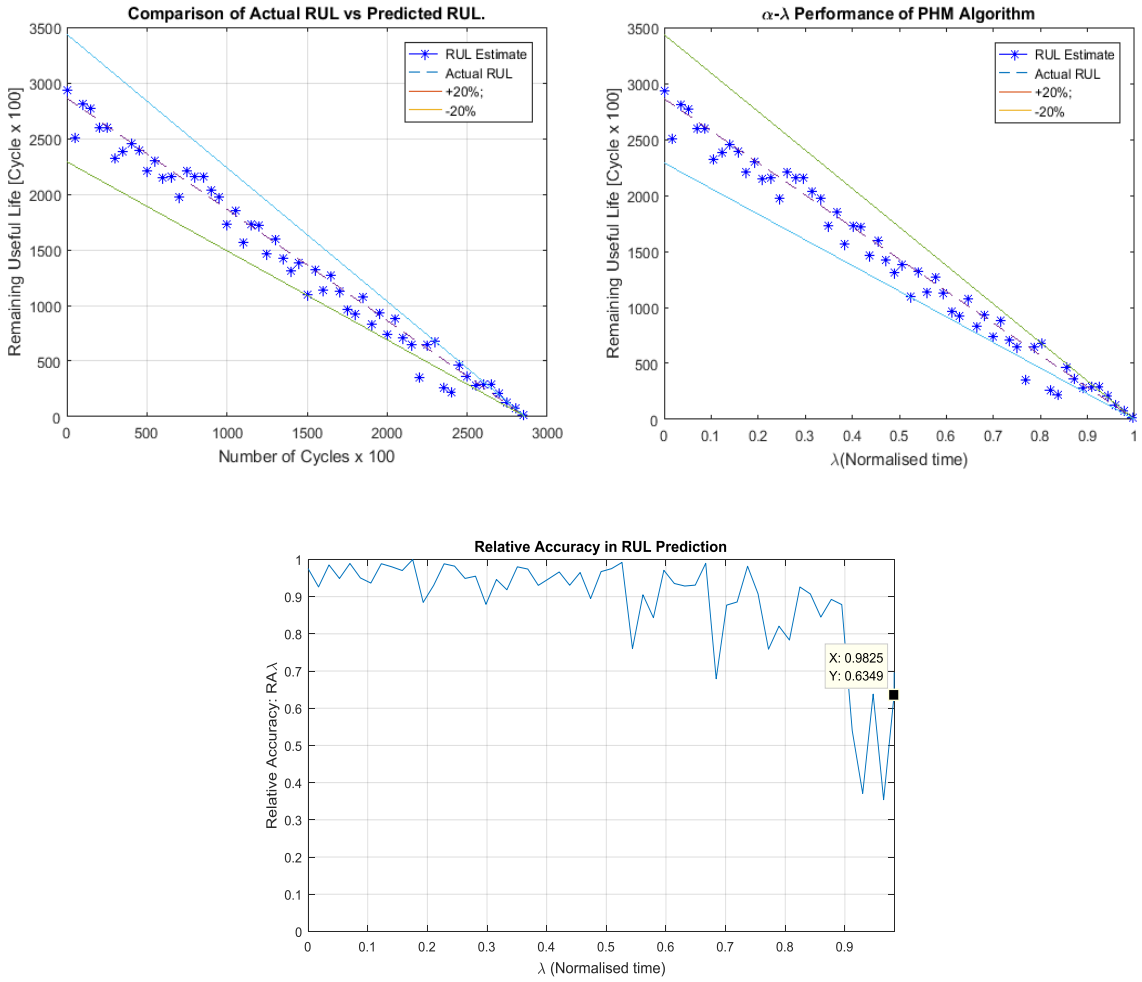
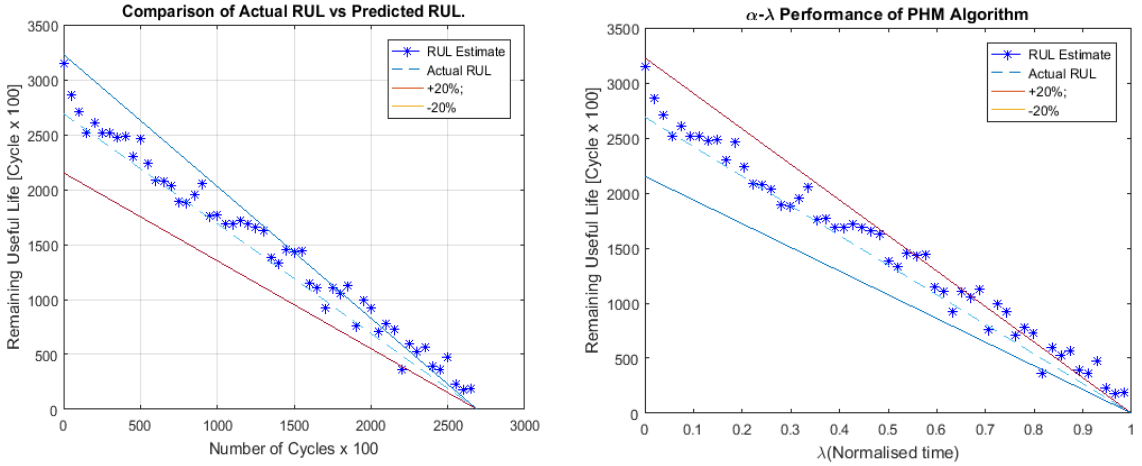


Figure 5.27. Sample 4. RUL, alpha-lambda and relative accuracy respectively, for sample 4.

Sample 5



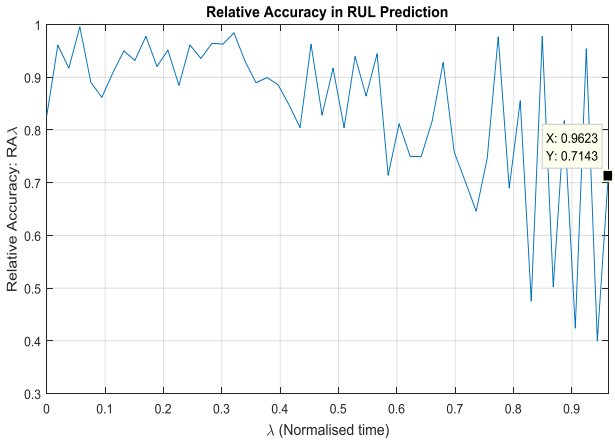
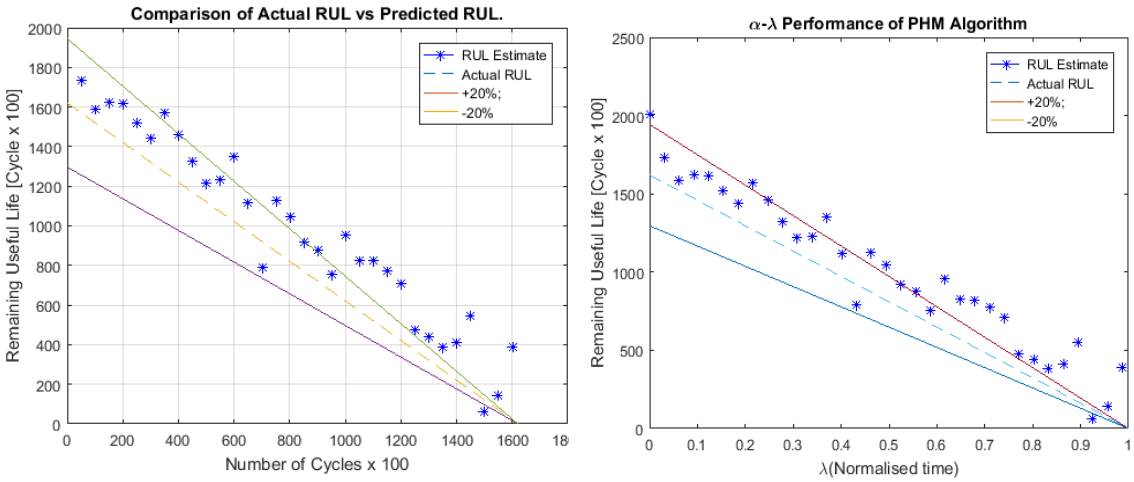


Figure 5.28. Sample 5. RUL, alpha-lambda and relative accuracy respectively, for sample 5.

Sample 6



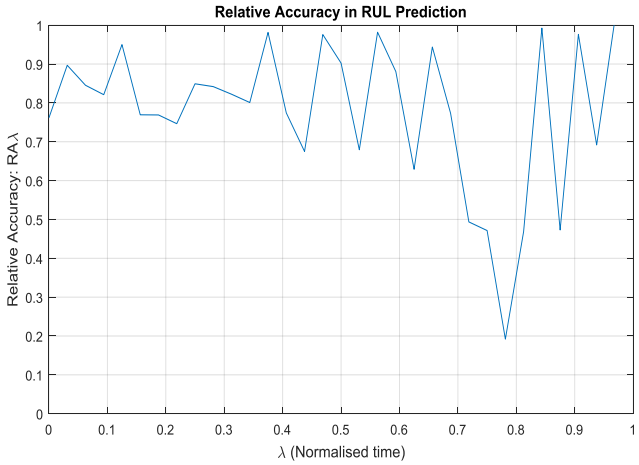


Figure 5.29. Sample 6. RUL, alpha-lambda and relative accuracy respectively, for sample 6.

Conclusion

This chapter has developed a simple physics based model based upon arc erosion of the contact and its effect on the contact resistance due to the reduction of contact surface area. A physics based model of the arcing and the erosion process has been developed to enable mass loss to be calculated, the results are used in a state space model. This model was then used with particle filtering to enable a prediction of the RUL to be estimated.

Although the model stands up to the prognostic metrics used to bench mark the effectiveness, it still has considerable shortfalls. Any alterations to the model material loss parameter need to be gained from the arc model and heat equation, hence the model is not a true physics based model in the sense that it may adapt to changes in conditions such as loading, environmental or material deviations. Other flaws include the considerable computation power needed when using the particle filter, to get a reasonable estimate, 3000 particles where used, this evoked a great deal of processing time even on a Core™ i5 machine. The average time of one iteration was taken from Matlab™'s timing function and was found to be 29.91mS. A trade-off between iteration time and accuracy can be made. Reducing the amount of particles reduces the iteration time but also the accuracy of the prediction, whereas increasing the particle count increases the accuracy, but at the cost of time.

This chapter provides the following conclusions and contributions

- The development of a model for the estimation of secondary parameters within the relay such as coil current, contact bounce and contact position.
- A model to enable the primary effects from arcing damage to be examined. The model incorporates the ability to look at heat flow through the contacts, electrical voltage across the contact due to heating and mass loss from the contact.
- The mass loss estimation which contributes towards the development of a state space physics of failure model.
- Estimation of RUL from this model.

Chapter 6 - Data Driven Model Parameter Identification

Introduction

The physics of failure model developed in Chapter 5, was unsatisfactory in a number of areas; the models material loss parameter needed to be gained from the arc model and heat equation and adaption to changes in conditions such as loading, environmental or material deviations where limited. As well as this, the computational power needed when using the particle filter was found to be excessive.

One of the original objectives of this research was to look into a methodology for on-board, real time prognostics, hence this chapter considers a different approach based upon the contact resistance failure data to develop the model in real time and then use this model to form a prognostic.

The literature review in Chapter 3 reports numerous data driven methodologies. The methods reported tended to fall into the classes of statistical or stochastic based models and artificial intelligence based. Both methodologies tend to require a considerable amount of data to either infer a model or learn patterns from. Due to the considerable time taken to fail a relay, even through the use of accelerated testing this still amounted to a period of around two weeks, masses of data were not available. Hence a new approach for devising the prognostic model in real time is proposed, using a sliding window and modelling methodology based upon system identification.

6.1 System Identification

From the work carried out in Chapter 5, an accurate, dynamic model based upon all the physical parameters involved is very difficult to infer. System identification is a technique which uses the input and output signals measured from a system to estimate the values of adjustable parameters in a given model structure. Obtaining a good model of the system depends on how well the measured data reflects the behaviour of the system.

System Identification requires a model structure. A model structure is a mathematical relationship between input and output variables that contains unknown parameters. Examples of model structures are transfer functions with adjustable poles and zeros, state space equations with unknown system matrices, and nonlinear parameterized functions (Ljung, 1999).

Black-box modelling is useful when the primary interest is in fitting the data regardless of a particular mathematical structure of the model. Black-box modelling is usually a trial-and-error process, where estimates of the parameters of various structures are produced and the results are compared. A simple linear model structure may first be tried, with progression to more complex structures as the need arises. The model structure may be chosen due to familiarity with this structure or because of specific application needs. Examples of Black Box modelling include ARMA, ARMAX, and ARIMA methods (Ljung, 1999).

An alternative method called Grey-Box modelling uses a model structure that can be deduced from physical principles. For example, the mathematical relationship between the input force $F(t)$ and the resulting mass displacement in the mass-spring-damper system is well known. In state-space form, the model is given by:

$$\frac{dx}{dt} = Ax(t) + BF(t)$$
$$y(t) = Cx(t)$$

where $x(t) = [y(t); V(t)]$ are the state vectors. The coefficients *A*, *B*, and *C* are functions of the model parameters.

Here, the model structure is fully known but the values of its parameters e.g. mass, damping and spring constant are unknown. In the grey-box approach, the data is used to estimate the values of the unknown parameters of the model structure.

In general, grey-box models are built by:

1. Creating a template model structure.
2. Configuring the model parameters with initial values and constraints (if any).
3. Applying an estimation method to the model structure and computing the model parameter values.

The process of model selection assumes that a series of experiments is designed so that a sufficient data set can be obtained to fit a model of the system. The classic system identification loop is depicted in Fig (6.1) (Ljung, 1999).

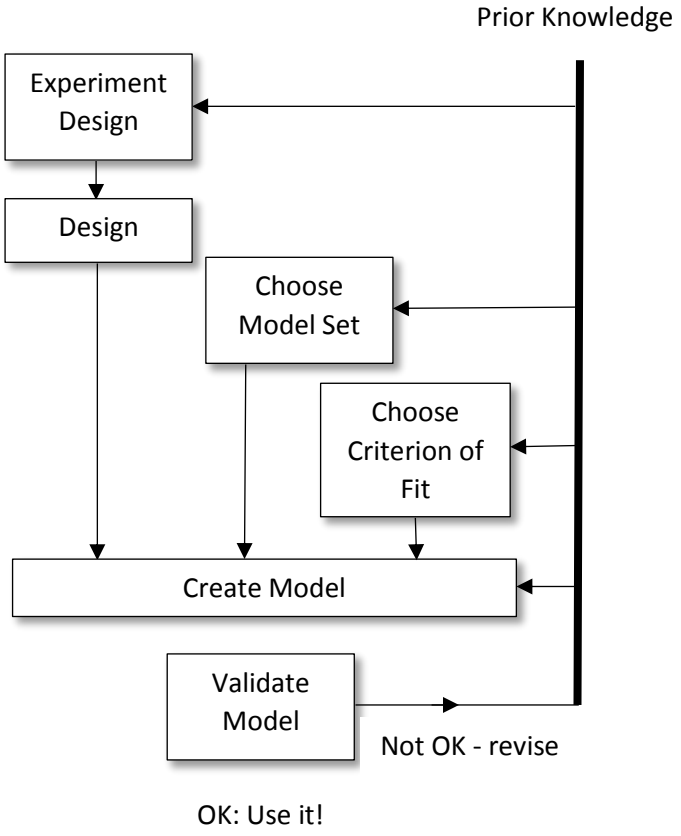


Figure 6.1. Depicting the classic system identification design loop.

6.1.2 Model Uncertainty

There will always be some uncertainty from estimating the model parameters from data and their nominal values will be accurate within a confidence region. The magnitude of this region is determined by the values of the parameter uncertainties computed during estimation. The magnitude of the uncertainties provide a measure of the reliability of the model. Large uncertainties in parameters can result from unnecessarily high model orders, inadequate excitation levels in the input data, and poor signal-to-noise ratio in measured data (Ljung, 1999).

6.1.3 Development of Models

From the literature review, various authors have proposed data driven approaches for relay reliability assessment. (Fang et al., 2006) studied traditional reliability assessment methods for electromagnetic relays that are based on censored failure time data; and concluded that they provided very little reliability information. (Qiong et al. & Xuerong et al., 2010) stated that in order to predict the life of the relay, a metric of degradation needed to be defined, methods explored include dynamic contact

resistance, pick-up time, over-travel time, the rebound duration, closing time, the fluctuation coefficient respectively as the predicted variables of the abrasion failure, bridging failure and the contamination failure (Qiong et al., 2010) showed by using time series analysis and by measuring characteristic parameters as predicted variables, the life of relay can be obtained. However, the conclusions showed the predicted accuracy is greatly influenced by the complex variations of characteristic parameters, and as a result it sometimes becomes too low to be accepted. Life prediction based on wavelet transform and ARMA (auto-regression moving average) time series was proposed to improve this (Yu, 2009).

A linear regression analysis method was proposed by (Xuerong et al., 2012) and has been used to establish the linear degradation model which regards the operation time as the independent variable and the predicted variables of the failure mechanisms as the dependent variable.

In this chapter, two novel methods based upon series identification methods will be looked at and evaluated in terms their effectiveness.

1) Grey System Models from time series data
2) Sliding Window Recursive Least Square, ARMA model.

6.2 Grey System Model for Time Series Prognostics

Grey theory (Deng, 1982) offers a truly multidisciplinary approach to dealing with systems that are characterized by poor information and/or where information is missing. Literature reports the theory's use in a multitude of fields, including data processing, modelling, system analysis, decision making and control. Potential attributes of Grey theory include the assumptions regarding the statistical distributions of the data are unnecessary, coupled with advantage of not having the cost/time in developing a physical model. Therefore given a time series relating to a failure parameter of a complex system that is particularly difficult to model or accurately predict, the Grey series becomes an attractive solution. It is for this reason that its feasibility for predicting the RUL is explored.

The measurement of contact resistance form a time series which is a collection of data points which are generally sampled equally in time intervals. Time series prediction refers to the process by which the future values of a system is forecast based on the information obtained from past and current data points. Generally, a pre-defined

mathematical model is used to make accurate predictions. Grey theory has become quite popular with its ability to deal with the systems that have partially unknown parameters. As an improvement on conventional statistical models, Grey models require only a limited amount of data to estimate the behaviour of unknown systems (Deng, 1989). The application of the grey prediction model was investigated for the failure prognostics of ball gate arrays in electronics by (Gu, Jie, et al., 2010). The grey prediction demonstrated a higher level of accuracy when dealing with small sample size of data.

6.2.1 Fundamental Principles of Grey System Model

GM(1,1) is a special case of GM(1,N) in which N=1, that is, the differential equation model of grey system theory of one variable of one order.

Establishment of GM(1,1) only requires a progression x(0), with a basic equation as follows:

GM(1,1) Model

$$x^{(0)} = x^{(0)}(1), x^{(0)}(2), \dots x^{(0)}(n) \tag{145}$$

AGO is the accumulated generating operation, and is given by

$$AGO, x^{(1)}(k) = \sum_{m=1}^{k} x^{(0)}(m) \tag{146}$$

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \dots x^{(1)}(n-1) + x^{(0)}(n)\right) \tag{147}$$

$$= \left(x^{(0)}(1), x^{(1)}(1) + x^{(0)}(2), \dots, x^{(1)}(n-1) + x^{(0)}(n)\right) \tag{148}$$

The GM(1,1) model uses a first-order differential equation.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \tag{149}$$

with the parameter series $\hat{a} = \left[\frac{a}{u}\right]$ and is given by a least squares estimation

$$\hat{a} = (B^T * B)^{-1} * B^T * y_n \tag{150}$$

where $B = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(1) + x^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(3)\right] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[x^{(1)}(n-1) + x^{(1)}(n)\right] & 1 \end{bmatrix}$ (151)

and $y^n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$ (152)

Solution of the differential equation is then:

$$x^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{u}{a}\right] e^{-ak} + \frac{u}{a} \quad (153)$$

Then, the Inverse Accumulated Generating Operation (IAGO) can be used to obtain the inverse data series from the AGO. It is then used to transform the forecasted AGO data back into the original time series and is achieved by the following equation:

$$x^{(0)}(k+1) = x^{(1)}(k+1) - X^{(1)}(k) \quad (154)$$

where $x^{(0)}$ is the predicted original series. Combining equations (153) and (154), we can get equation (155) as follows:

$$x^{(0)}(k+1) = \left(x^{(0)}(1) - \frac{u}{a}\right)\left(e^{-ak} - e^{-(k-1)}\right) \quad (155)$$

In order to optimise the processing of the data, which may be in the order of hundreds of thousands of samples and may take a considerable time to process to find the best-fit parameters, a sliding window approach is employed.

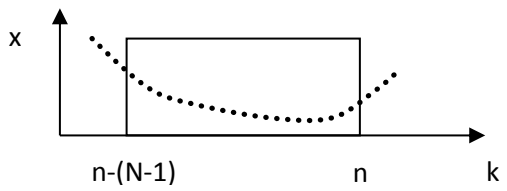


Figure 6.2. Sliding window implementation.

Damage may not be significant or vary very slowly, hence resulting in very small changes. However, at some point, the damage will start to change quickly, hence producing a steeper slope within the model. The other advantage of this method is the potential sensitivity to small parameter changes in the contact resistance as they

develop due to dynamic factors like temperature, variation in loading (reactive parameters), environment, vibration, voltage and current.

The results are based on two sets of contact resistance data from the experimental work carried out in Chapter 4. Figures 6.3 and 6.4. The threshold was calculated for both data sets based upon the welding threshold described in Chapter 4. It can be seen that the prediction of RUL for the first dataset in Figure 6.3 set is a good approximation of where it failed. This is in turn due to the steady upward increase in the contact resistance. However, in the data from Figure 6.4, the RUL is drastically cut short due to high level of overshoot within the data, which one may expect intuitively to cause a failure.

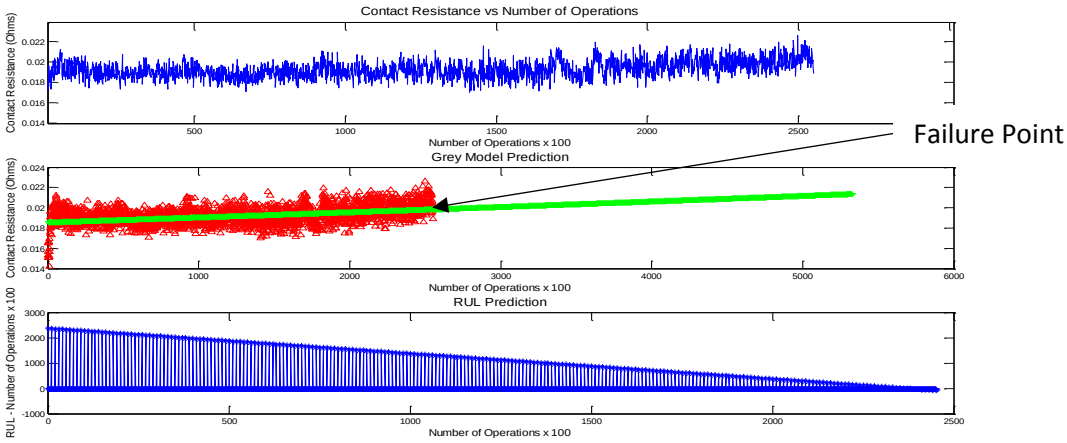


Figure 6.3. RUL-based upon steady contact resistance increase.

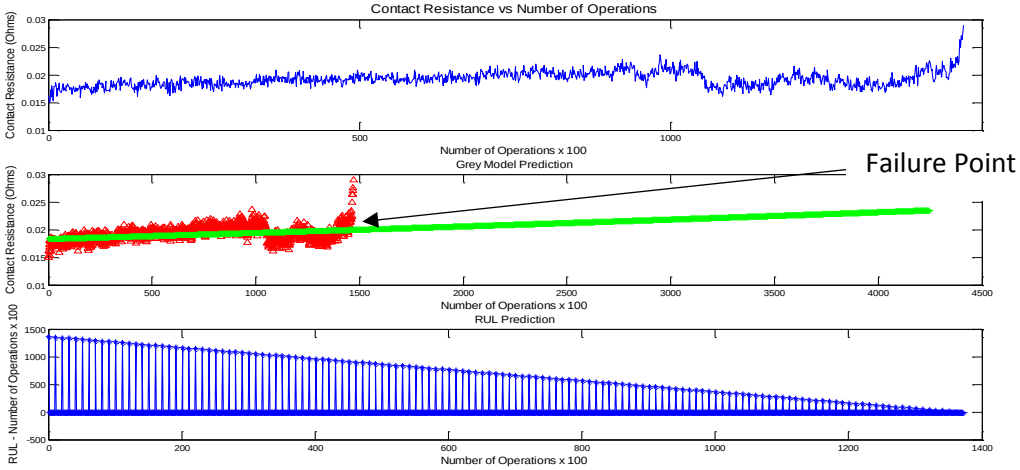


Figure 6.4. RUL prediction based upon high levels of overshoot in the contact resistance

The Grey Model approach was found to be suitable for estimating systems that have partially unknown parameters, but has downfalls in that the correction terms are not estimated in a learning paradigm to improve the accuracy and precision of the algorithm for long-term prediction. The way a fault or fault indicator evolves over time entails a large amount of uncertainty. To enable the accurate and precise prediction of the time to failure for the relay, the critical degradation variable such as the contact resistance must be considered as a random variable with an associated probability distribution vector. By obtaining the probability distribution of the failure, confidence intervals and other important attributes may be computed. This however is not obtainable with the Grey model.

Bayesian estimation techniques employ a state dynamic model and a measurement model to predict the posterior probability density function of the state, that is, to predict the time evolution of a fault or fatigue damage and are finding application domains in machinery fault diagnosis and prognosis of the remaining useful life of a failing component/subsystem (Orchard, 2009). The work done in Chapter 5 has already made use of such a technique, allowing a prognosis to be based on recursive Bayesian estimation technique that used information from a physic based model. Prognosis or long-term prediction for the failure evolution is based on both an accurate estimation of the current state and a model describing the fault progression.

If the incipient failure is detected and isolated at the early stages of the fault initiation, it is reasonable to assume that sensor data will be available for a certain time window allowing for corrective measures to be taken, that is, improvements in model parameter estimates so that prognosis will provide accurate and precise prediction of the time to failure. At the end of the observation window, the prediction outcome is passed on to the user (operator, maintainer), and additional adjustments are not feasible because corrective action must be taken to avoid a catastrophic event (Orchard, Wu and Vachtsevanos, 2005).

Figure 6.5 below shows a model-based prognostic scheme. Input from the diagnostic block is combined with stress profiles and feeds into the fault growth model. An estimation method (such as Kalman or Particle filtering) is called upon to propagate the fault model initially, one step at a time, while model parameters are updated online in real time as new sensor data become available (Vachtsevanos, 2006).

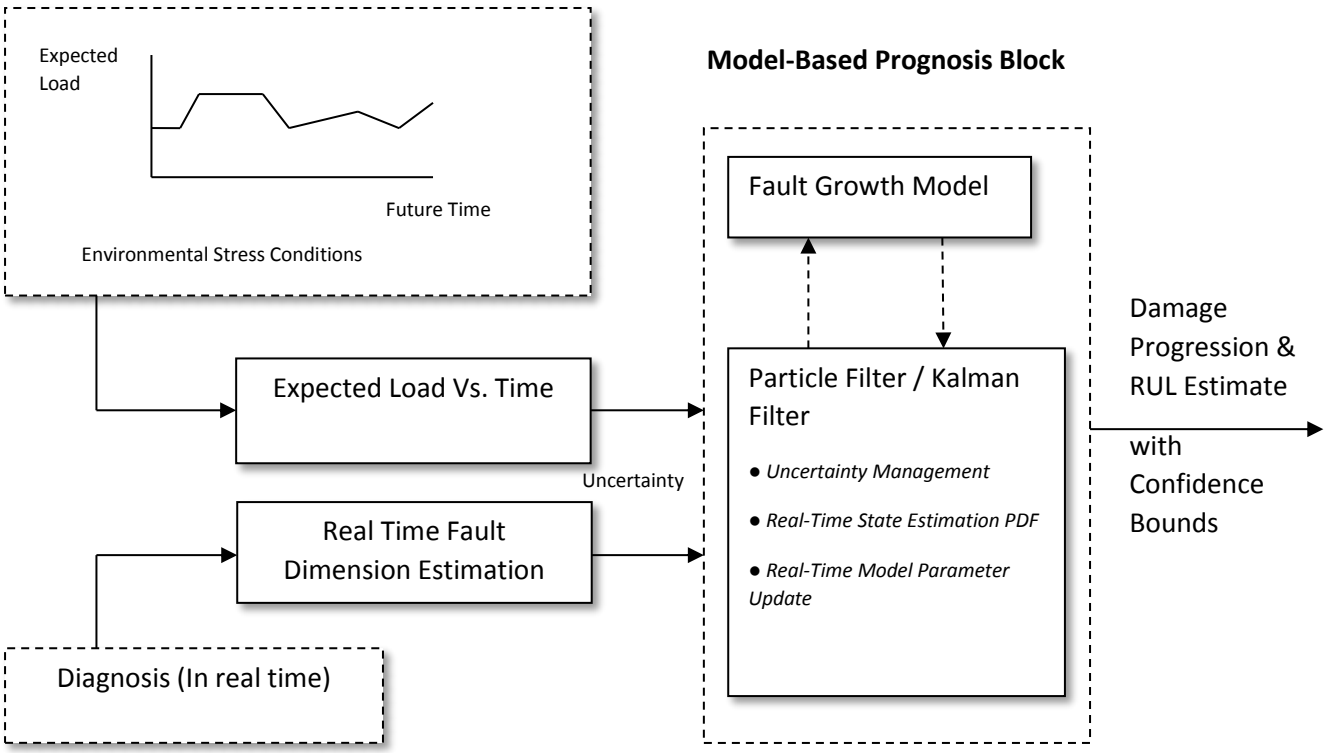


Figure 6.5. A model-based prognostic scheme.

Eventually, the model is allowed to perform long-term prognosis of the remaining useful life of the failing component/system with confidence bounds. The fault model PDF is convolved with the hazard zone PDF when the former reaches the threshold bounds and the resultant PDF is projected along the time axis (which is usually measured in "cycles" of operation) depicting the system's remaining life statistics.

As seen in Chapter 3, physics-based models provide a means to calculate the damage to critical components as a function of operating conditions and assess the cumulative effects in terms of component life usage. However, an accurate physics based model of the process is often difficult to achieve. In the next section the physics based model will be replaced with an ARMA based state model whose coefficients are estimated from the degradation data by a Recursive Least Square (RLS) sliding window algorithm. The ARMA state model will then be integrated with stochastic modelling techniques, the model can be used to evaluate the distribution of remaining useful life as a function of uncertainties in component strength/stress properties and loading conditions for a particular fault. Bayesian estimation techniques will be used to satisfy these requirements in the form of Kalman filtering. In the framework proposed in Figure 3.3, an ARMA model of the feature vector, replaces the traditional empirical or physics-based degradation models. This is used in conjunction with the Kalman filter

and measurement data to enable an estimation of the feature vector to be made up to the failure threshold.

Algorithm: SWRLS Parameter Estimation, Filtering and RUL prediction

For n < 0: *h = 0,* $\widehat{\boldsymbol{\theta}}_{-1} = 0, \overline{\boldsymbol{\theta}}_{-1} = 0, \boldsymbol{P}_{-1} = \frac{1}{\delta} I_N, \overline{\boldsymbol{P}}_{-1} = \frac{1}{\delta} I_N$

For data sequence at n = 0,1,…

1. Estimation of parameters from new data at nth instant

$$\overline{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_{n-1} + \boldsymbol{K}_n\{ y_n - h_n^T \widehat{\boldsymbol{\theta}}_{n-1}\}$$

2. Calculate Kalman gain

$$\boldsymbol{K}_n = \frac{\boldsymbol{P_{n-1}} h_n}{1 + h_n^T \boldsymbol{P_{n-1}} h_n}$$

3. Calculate covariance

$$\overline{\boldsymbol{P}}_n = \boldsymbol{P_{n-1}} - \boldsymbol{K}_n \, h_n^T \boldsymbol{P}_{n-1}$$

For

$$U_n \leq U_{\boldsymbol{threshold}}$$

Where $U_{\boldsymbol{threshold}}$ is the point of where the component is deemed to start failing

then 13.

Else

The Kalman filtering algorithm consists of the following steps:

4. Initialize variables at time step

$\boldsymbol{P}_0 = \boldsymbol{P}_{x0}, \hat{x}_0 = \bar{x}_0;.$

For k = n, k is fixed and the prediction N is projected e.g. N = 1,2,3,…to threshold U

5. Project state at the next time step, where A & B are the $\hat{\theta}$ *estimates and u is taken at n + L,*

$\widehat{\boldsymbol{X}}_{(k+N|k)} = \boldsymbol{A}_{(k+N|k)} \widehat{\boldsymbol{X}}_{(k|k)} + \boldsymbol{B}_k u_k$

6.Take measurement, $z_k = \boldsymbol{H} x_k + v_k$

7. Calculate error covariance before update, $\boldsymbol{P}_{(k+N|k)} = \boldsymbol{A}_{(k+N,k)} \boldsymbol{P}_{(k|k)} \boldsymbol{A}^T_{(k+N,k)} + \boldsymbol{Q}_{(k+N,k)}$

8. Calculate the Kalman gain, $\boldsymbol{K}_k = \boldsymbol{P}_k \boldsymbol{H}_k^T \boldsymbol{R}_k^{-1} = \boldsymbol{P}_k^- \boldsymbol{H}_k^T (\boldsymbol{H}_k \boldsymbol{P}_k^- \boldsymbol{H}_k^T + \boldsymbol{R}_k)^{-1}$

9. Update estimate with measurement, $\hat{x}_{k+1} = \hat{x}_{k+1}^- + \boldsymbol{K}_{k+1}(z_{k+1} - \boldsymbol{H}_{k+1} \hat{x}_{k+1}^-)$

10. Calculate error covariance after measurement update, $\boldsymbol{P}_{k+1} + [(\boldsymbol{P}_{k+1}^-)^{-1} + \boldsymbol{H}_{k+1}^T \boldsymbol{R}_{k+1}^{-1} \boldsymbol{H}_{k+1}]^{-1}$

11. Repeat 5 to 10 until extrapolated to threshold value,

12. Report predicted RUL (and uncertainty).

13. Removal of data from window received at $(n-L+\delta)^{th}$ *instant*

$$\widehat{\boldsymbol{\theta}}_n = \overline{\boldsymbol{\theta}}_n + \boldsymbol{K}_n\{ y_{n-L+\delta} \, h_{n-L+\delta}^T \overline{\boldsymbol{\theta}}_{n-1}\}$$

14. Calculation of Kalman gain at n-L+δ *where* δ *is step size.*

$$\boldsymbol{K}_n = \frac{-\overline{\boldsymbol{P}}_n h_{n-L+\delta}}{1 + h_{n-L+\delta}^T \overline{\boldsymbol{P}}_n h_{n-L+\delta}}$$

15. Recalculation of Covariance

$$\boldsymbol{P}_n = \overline{\boldsymbol{P}}_n - \boldsymbol{K}_n h_{n-L+\delta}^T \overline{\boldsymbol{P}}_n$$

16. Return to step 1.

The complete algorithm is documented above and explained here. The initial parameters are set up for n < 0 and in 1 to 3, the RLS sliding window is set up and used in the estimation of parameters and ARMA state model formation from the new data at n^{th} instant. Next, because a prediction of RUL may not be relevant until a threshold is recognised, then the option to go from condition monitoring to prognosis is provided. Once the threshold of impending failure is met, then the Kalman filter is used in 4 to 10 to extrapolate the ARMA state model until the threshold of failure is met. From this a prediction of the RUL is given from the current measurement to the threshold. Lastly, is the removal of data from the window received at $(n-L+\delta)^{th}$ instant and the recalculation of the Kalman gain as well as the covariance at n-L+δ where δ is the step size. The processes is then started again, with the next data measurement.

One of the main advantages of this method apart from not having to have a physical model of the degradation processes is the noted computationally efficiency compared with that of the particle filter. Also, due to being able to quantify the model in terms of its fit to the data with an error measurement gives rise to the potential development of an on-line prognostic metric.

6.3 Kalman Filter for Prediction

In Chapter 3, the general operation of the Kalman Filter was outlined, based upon the justification of the process noise being additive white Gaussian for the w_k sequence on the process model. The Kalman Filter was used in this context as a one step ahead prediction, however for prognostics, a N step ahead prediction of the current measurement is required and this is can be implemented using the same identical methodology.

The equations (35) & (36) for N-step ahead prediction are modified to be:

$$\widehat{\boldsymbol{X}}_{(k+N|k)} = \boldsymbol{A}_{(k+N|k)} \widehat{\boldsymbol{X}}_{(k|k)} + \boldsymbol{B}_k u_k$$

$$\boldsymbol{P}_{(k+N|k)} = \boldsymbol{A}_{(k+N,k)} \boldsymbol{P}_{(k|k)} \boldsymbol{A}_{(k+N,k)}^T + \boldsymbol{Q}_{(k+N,k)}$$

Where

$\widehat{\boldsymbol{X}}_{(k|k)}$ = updated filter estimate at time t_k

$\widehat{\boldsymbol{X}}_{(k+N|k)}$ = predictive estimate of $\boldsymbol{X}$ at time t_{k+N} given all the measurement through t_k

$\boldsymbol{P}_{(k|k)}$ = error covariance associated with the filter estimate $\widehat{\boldsymbol{X}}_k^-$

$\boldsymbol{P}_{(k+N|k)}$ = error covariance associated with the predicted estimate

$\boldsymbol{A}_{(k+N,k)}$ = transition matrix from step *k* to *k* + *N*

$\boldsymbol{Q}_{(k+N,k)}$ = covariance of the cumulative effect of white noise inputs from step *k* to step *k* + *N*

From the above, there are two scenarios that may be considered. Firstly, N is fixed and k is allowed to evolve in integer time steps similar to a filter problem. The predictor may be thought of as an external element, in this case to the filter loop that is updated. For off-line analysis work, this may be used as a tool for giving a measure of the quality of the predictive state estimate by using the major diagonal of $\boldsymbol{P}_{(k+N|k)}$. Where online prediction is required it is $\widehat{\boldsymbol{X}}_{(k+N|k)}$ that becomes of interest and it should be noted that it is **not** necessary to calculate $\boldsymbol{P}_{(k+N|k)}$ in order to get an estimate of $\widehat{\boldsymbol{X}}_{(k+N|k)}$.

Secondly, *k* may be fixed, and then $\widehat{\boldsymbol{X}}_{(k+N|k)}$can be iterated along with its error covariance for 1 to N steps ahead utilising the Kalman gain and current measurement. In this case, the error covariance becomes extremely useful, especially for prognostics, as it offers a metric that allows the quality of the predictive estimate to be assessed, as the prediction is made up to the threshold of failure. It is this second scenario that is of interest in this Thesis.

Although the Kalman filter offers the optimum solution to state estimation for a linear model in additive Gaussian white noise, it is only as good as the information that is provided. The characteristics of the measurement noise and model are not evaluated as the Kalman filter operates and therefore it is has no way of knowing what is really the correct signal and what is unwanted noise. The only information the Kalman filter has on the noise processes is that which is specified in the ***Q*** and ***R*** covariance matrices which must be specified in advance. These matrices completely determine the action of the Kalman gain matrix and the final estimation of

the error covariance matrices, therefore as the component wears over time the characteristics of the noise signals may change and these will not be altered unless new values are assigned to the ***Q*** and ***R*** matrices.
As well as the filters performance in terms of the estimation accuracy being heavily reliant upon ***Q*** and ***R***, the state matrix ***A,*** input matrix ***B*** and ***H*** giving the ideal (noiseless) connection between the measurement and the state vector also needs to be accurate.
The covariance matrix ***R*** has been already been quantified in Chapter 4 for the measurement noise ***v***, and will consist of the uncertainty value which was found to consist mainly of normally distributed quantisation and flicker noise.
The process noise ***w***, however is harder to quantify, although it assumed to be Gaussian white noise, in reality it should cover all disturbances including; steps, spikes, coloured noise and modelling errors as well as external factors representative of environmental effects. From the ARMA model, one parameter that may be measured is the model accuracy, however other parameters can only be intelligently estimated. The noise elements of the matrix ***Q***, again like ***R***, are assumed to be uncorrelated and form the diagonal.

In reality, often the only way of setting the values of ***Q*** is by assigning variables at random and then 'tuning' them to reach the optimum performance in simulation studies. This can becomes increasingly time consuming when higher order systems are invoked containing many disturbance signals. The Kalman filter itself may also be of help in the 'tuning' of ***Q*** and ***R*** by observing the characteristics of the outputs. If the elements of ***Q*** are decreased in magnitude, or the elements of ***R*** increased in magnitude, this implies than there is relatively more noise on the measured signal than the states. This causes the Kalman filter to make the assumption that the prediction made by the state estimates from the model, are more reliable than the measurements, thus an emphasis is placed on prediction and less so on the measurement.

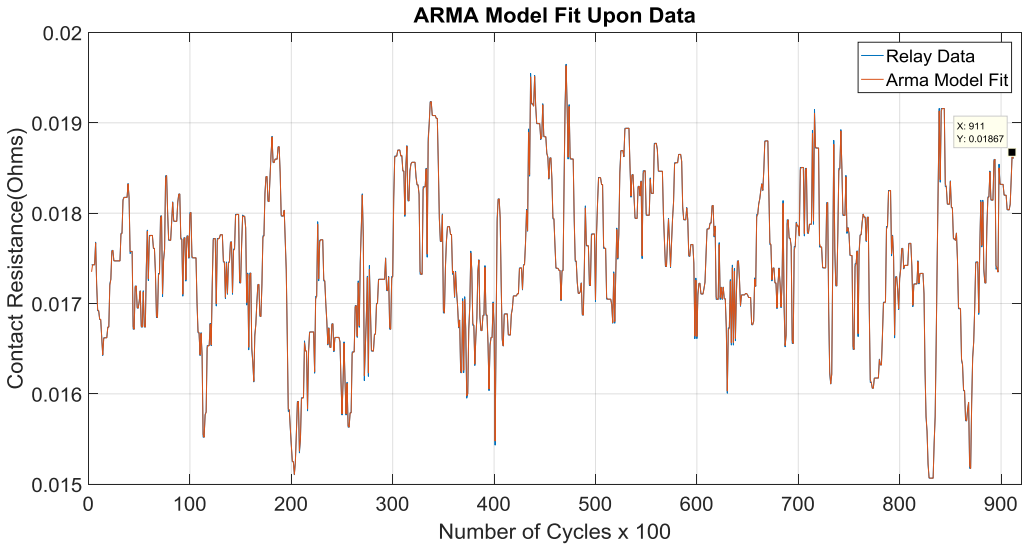
Conversely, if the elements of the ***Q*** matrix are increased in magnitude or the elements of ***R*** decreased, this implies there is less noise on the measured signal relative to the state model. The assumption is then made that the generated state estimate is less reliable than that of the measurement, thus requiring more

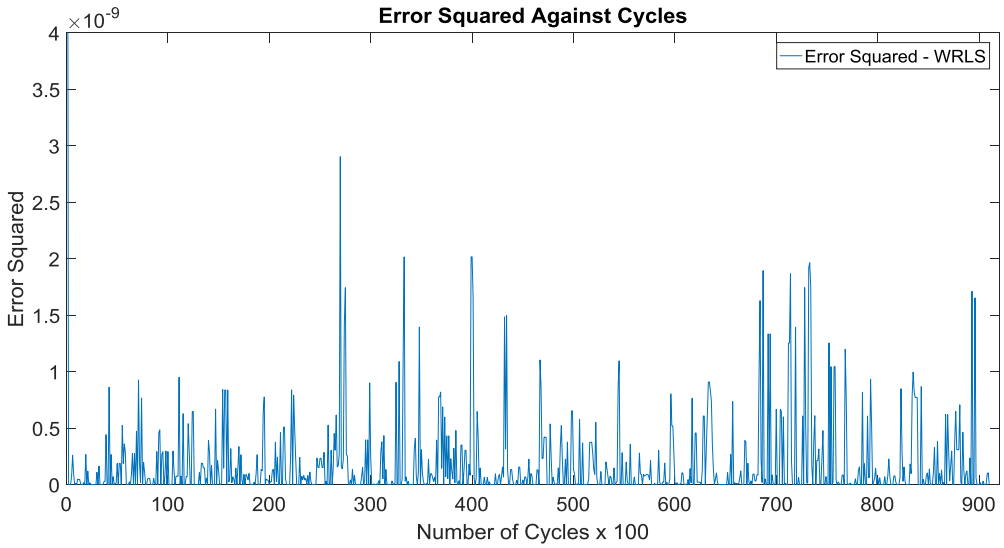
correction, causing the magnitude of the elements in the Kalman gain matrix to increase accordingly.

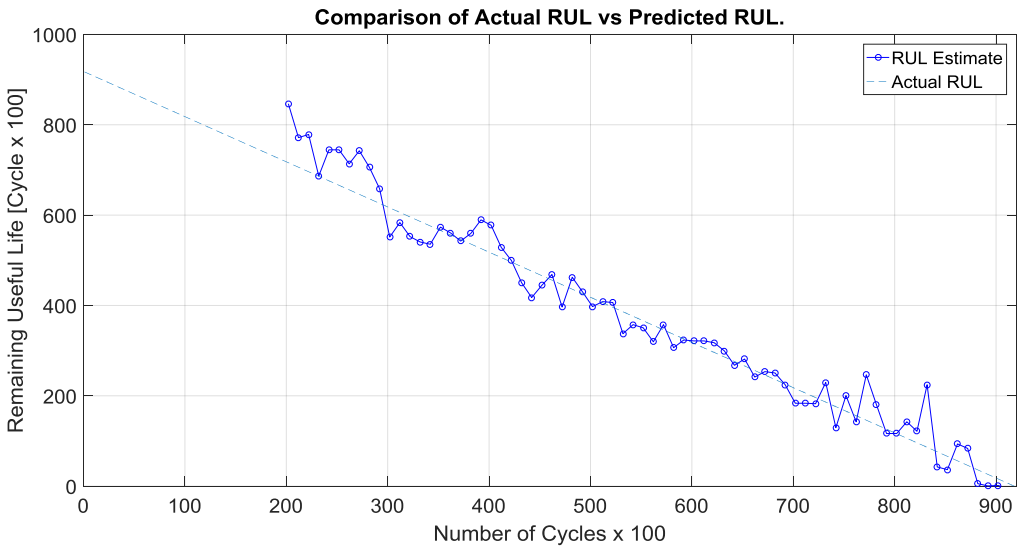
The final estimated error in the covariance matrix ***P***, should be small and relatively symmetrical, in general, small elements in ***P*** implies that Kalman filter has a high degree of trust in the prediction, conversely, high ***P*** elements indicate distrust.

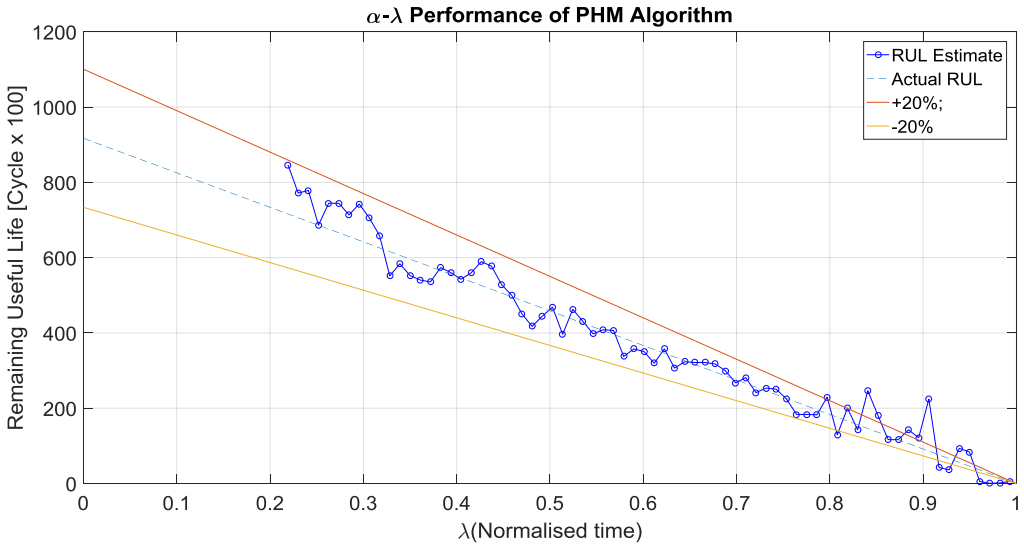
6.4 Results from SWRLS ARMA Model

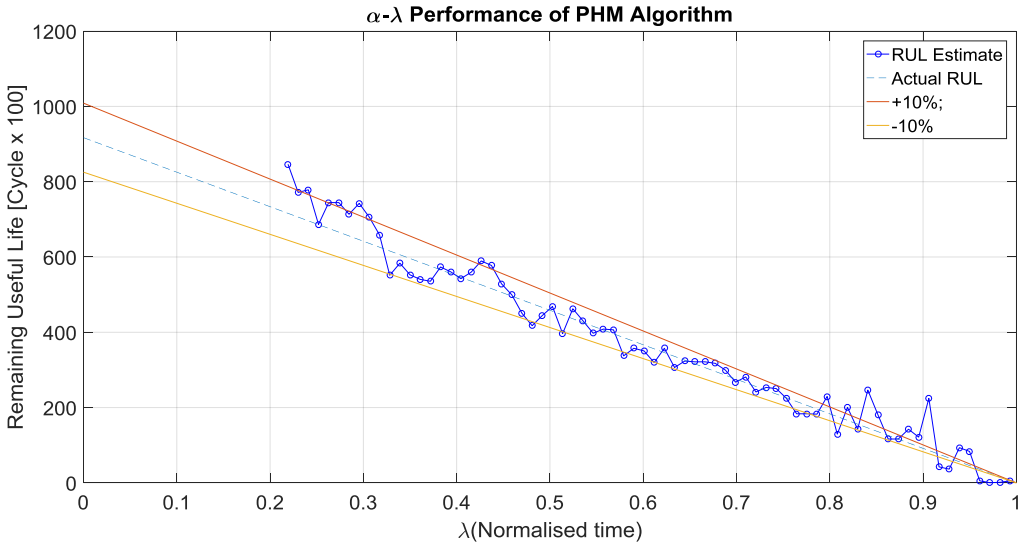
Sample 1 – 200

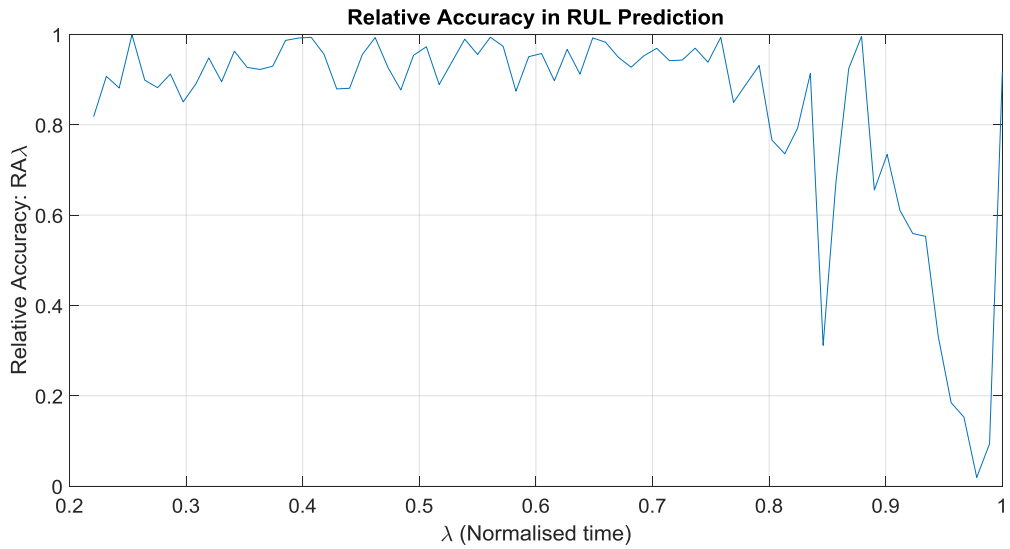





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000

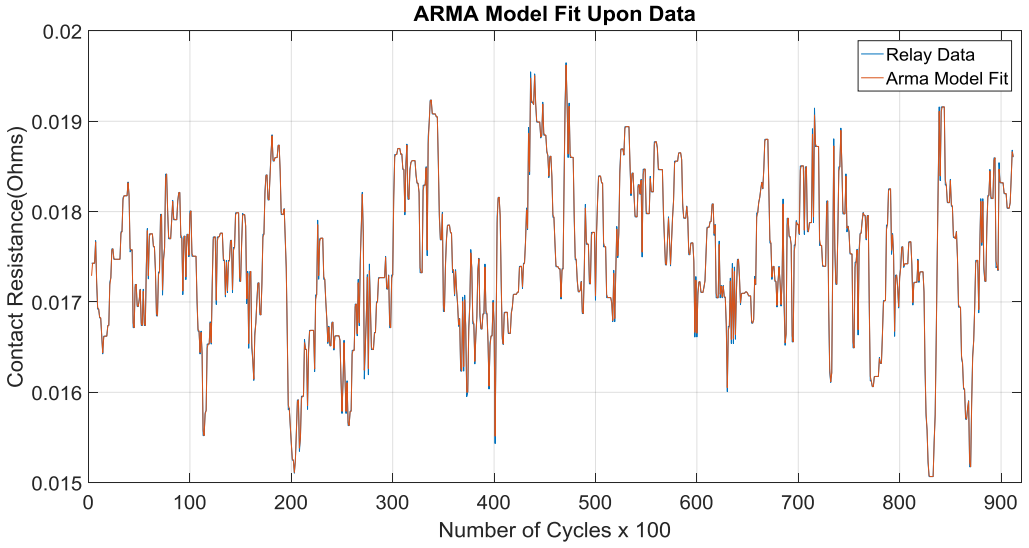

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

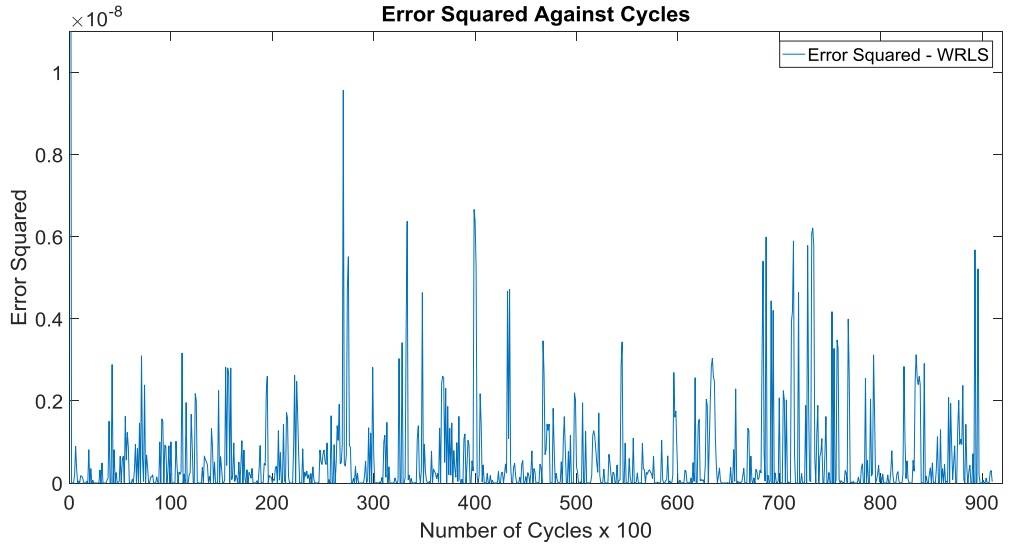

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

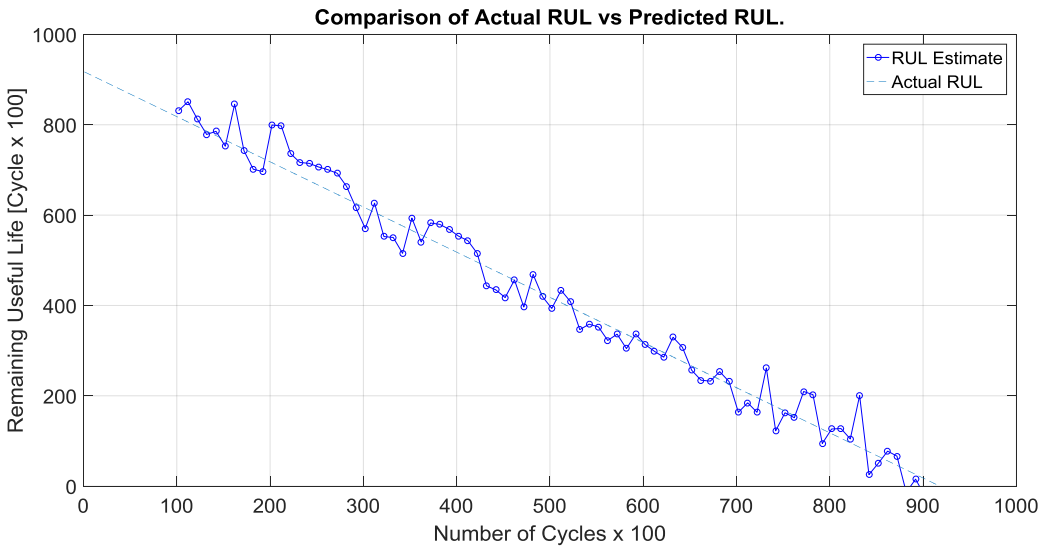


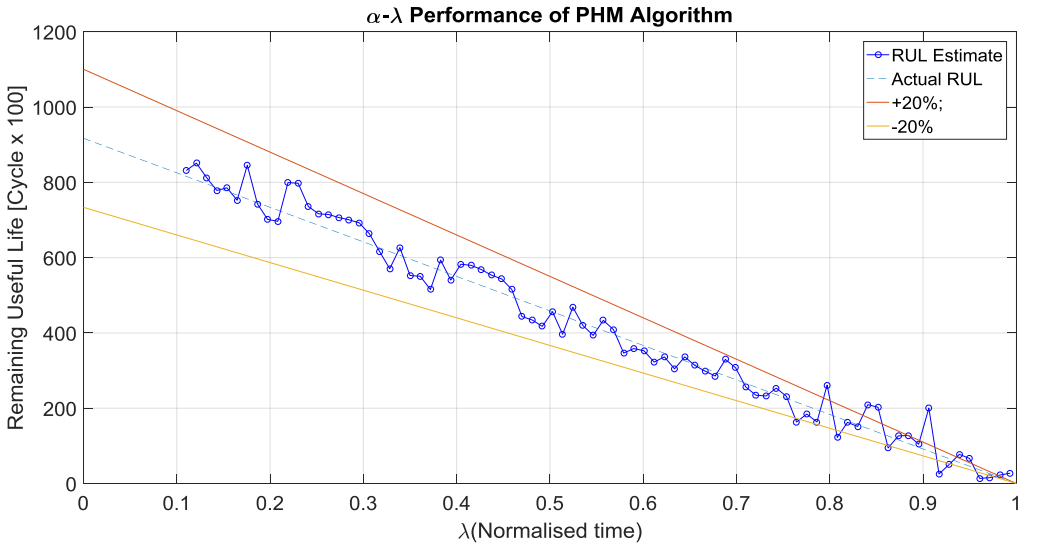
Figures 6.6. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 1 with 200 window.

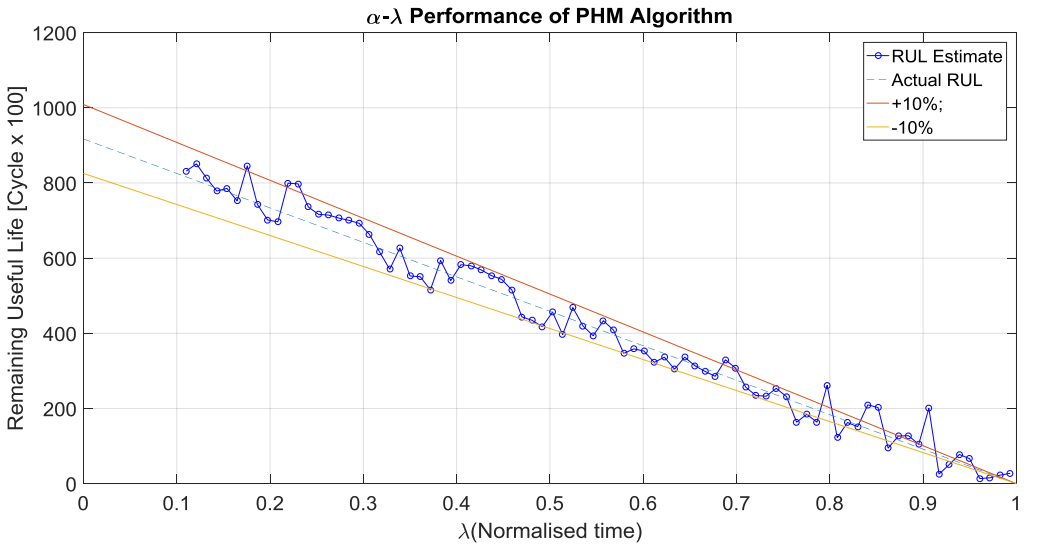
Sample 1 – 100

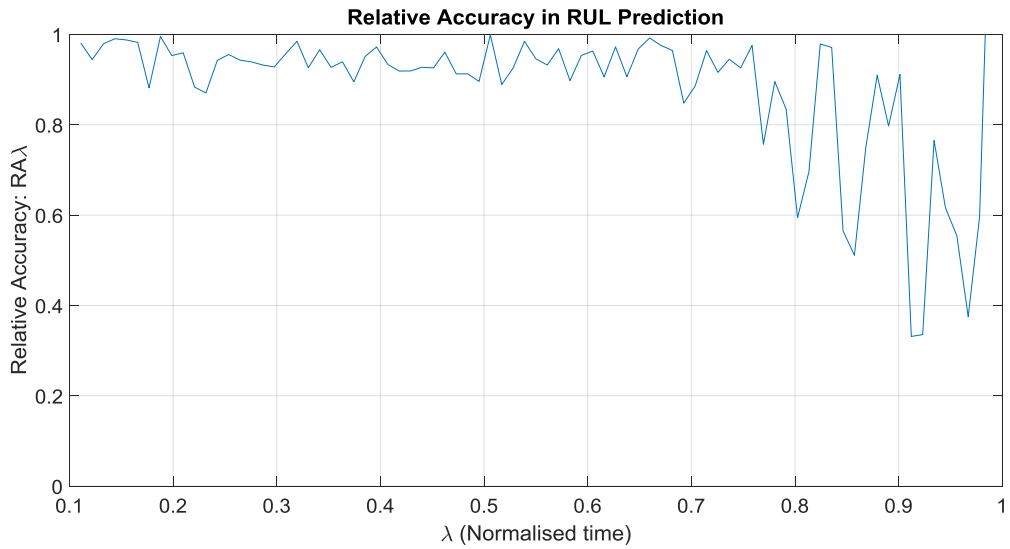





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000

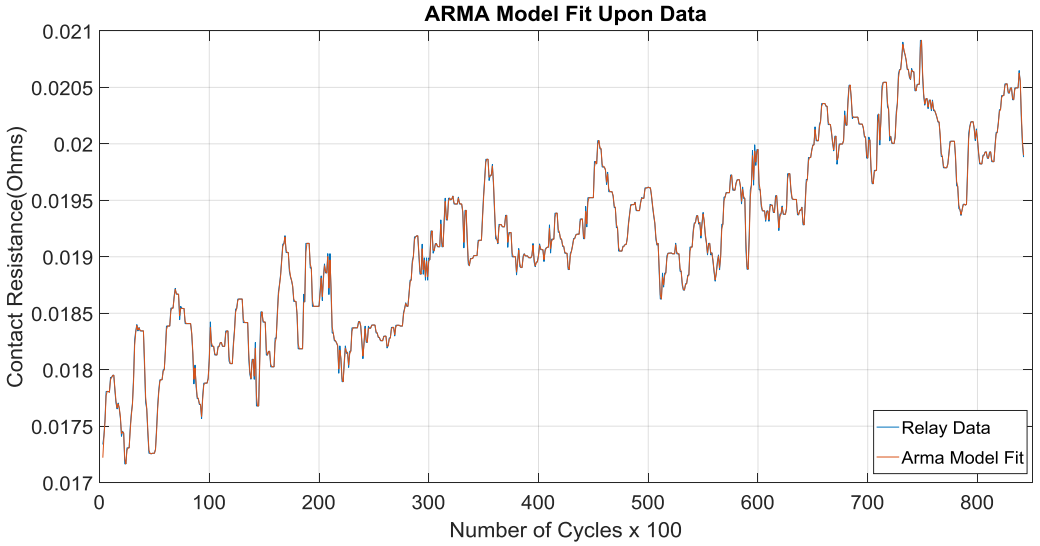

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

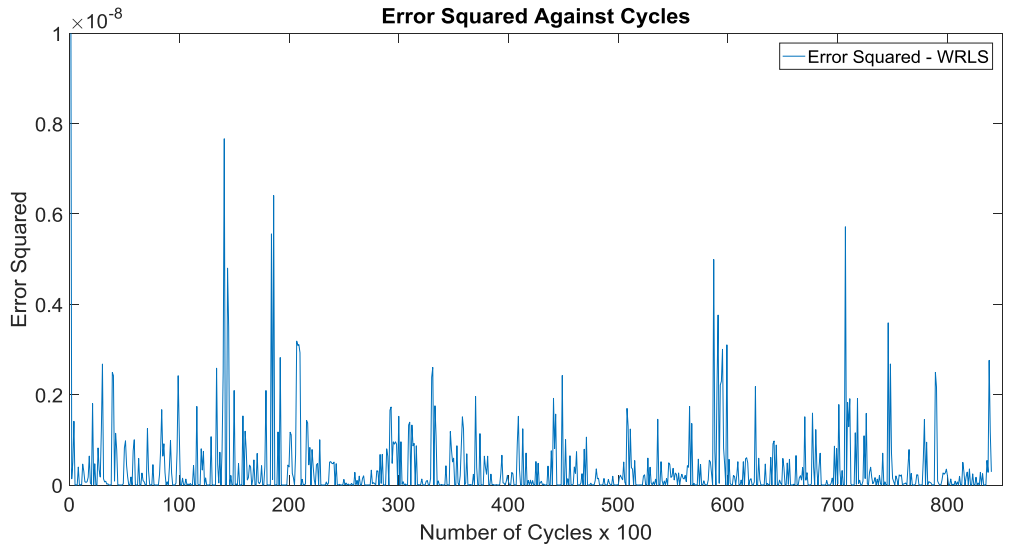

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

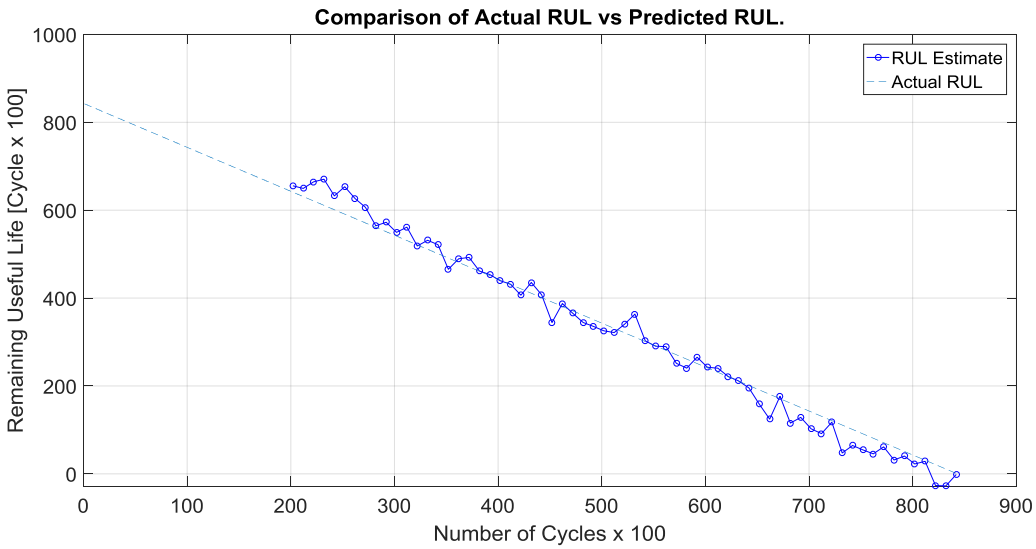


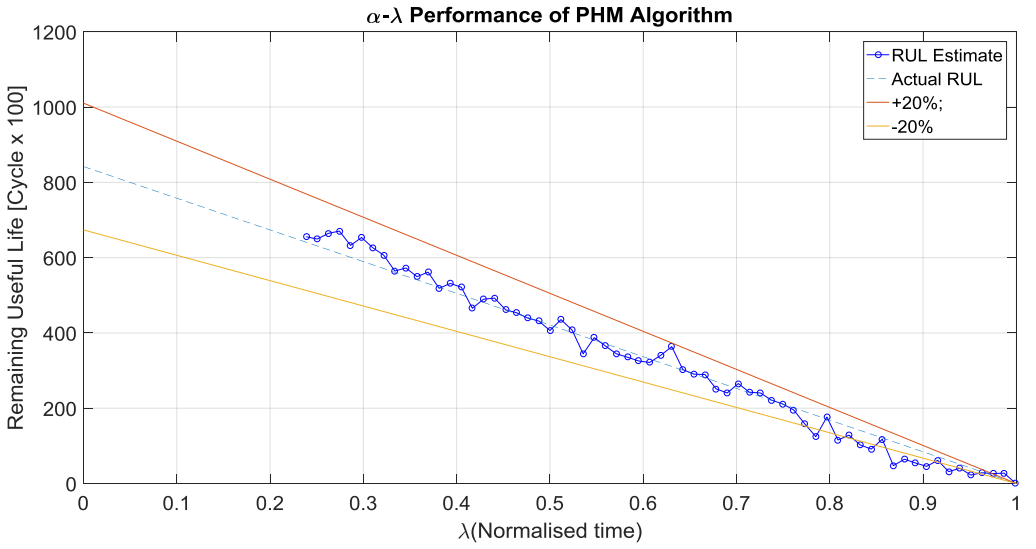
Figures 6.7. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 1 with 100 window.

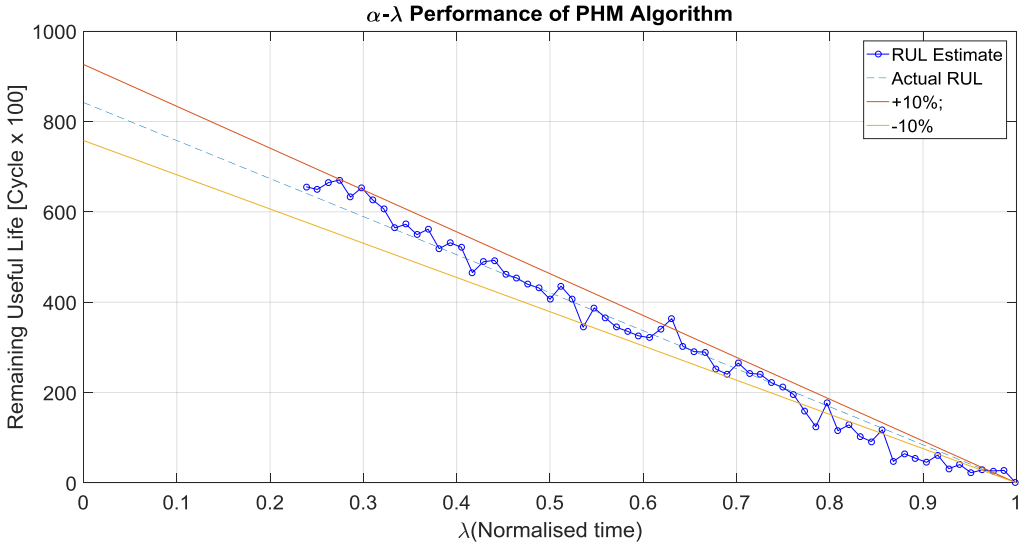
Sample 2 - 200

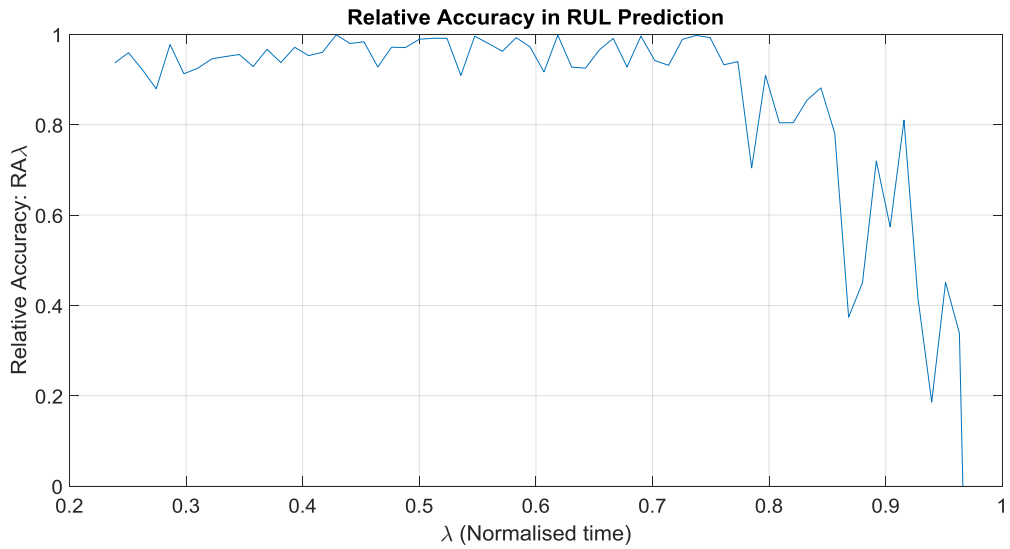





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000

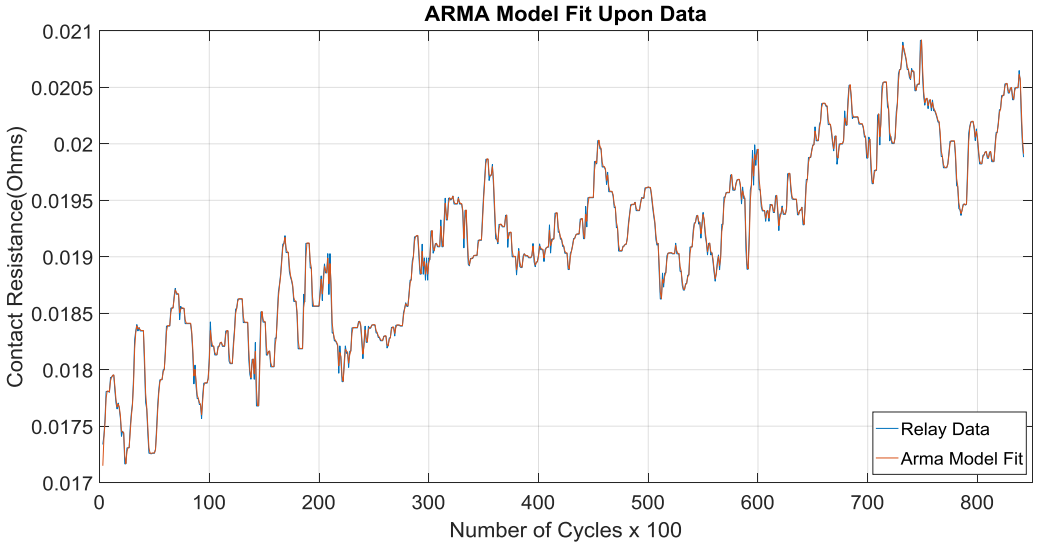

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

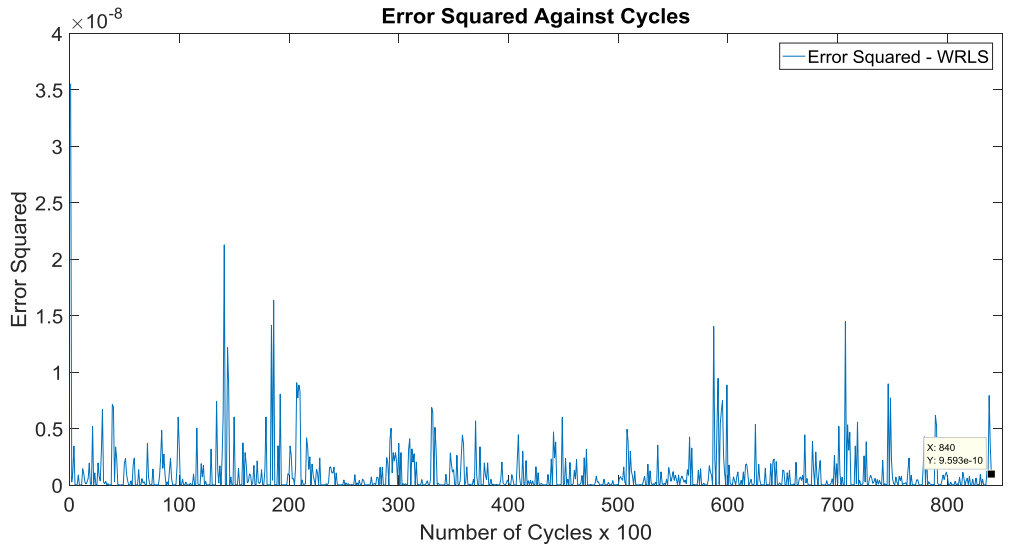

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1000

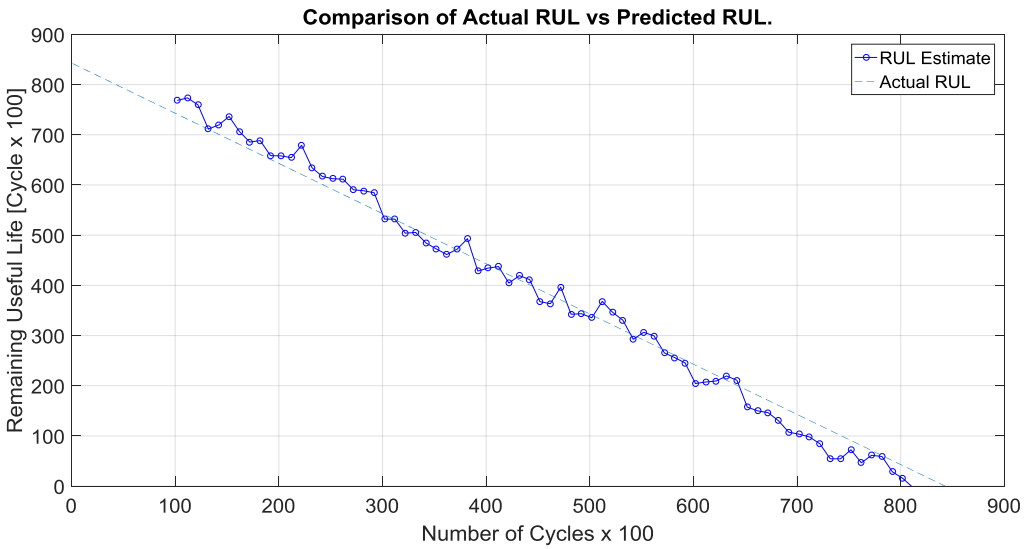


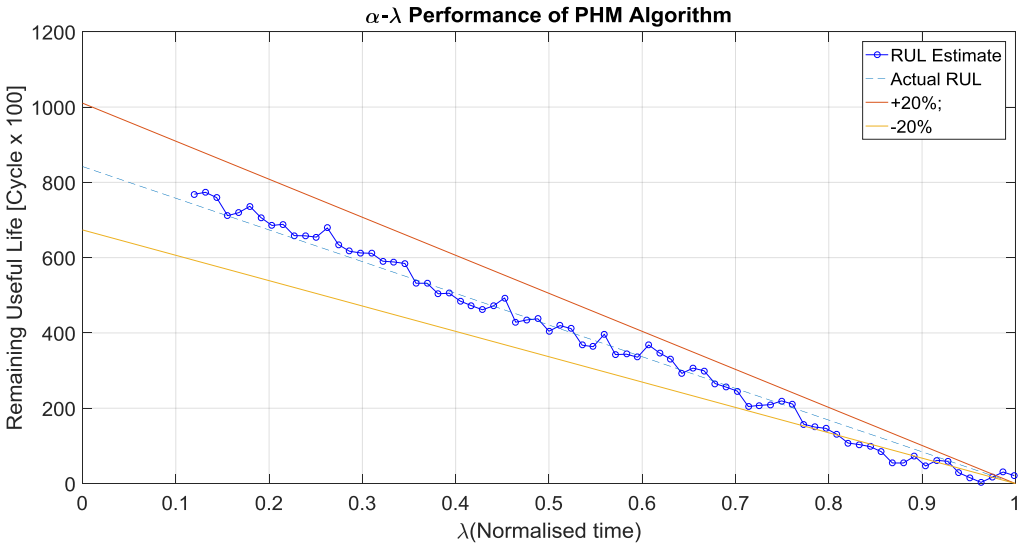
Figures 6.8. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 2 with 200 window.

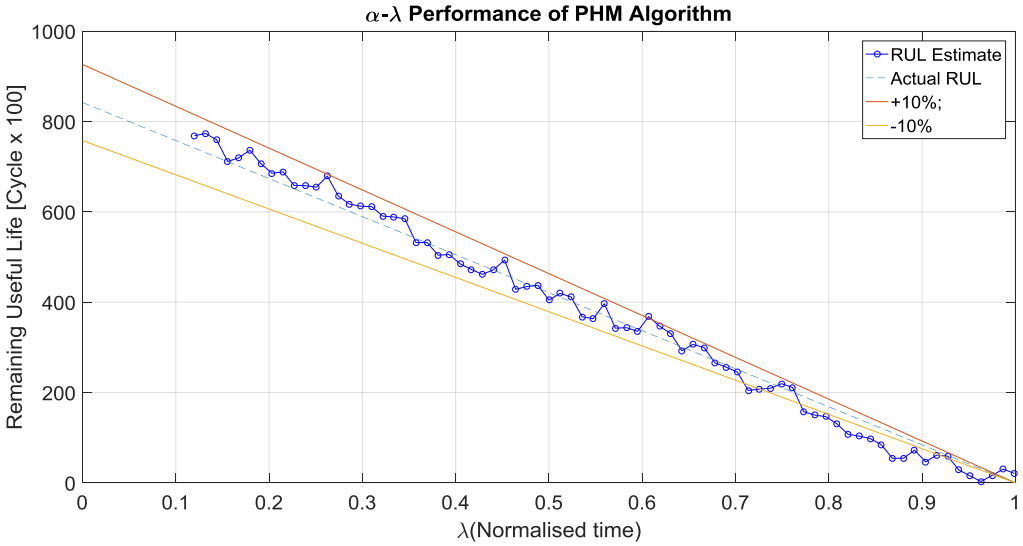
Sample 2 -100

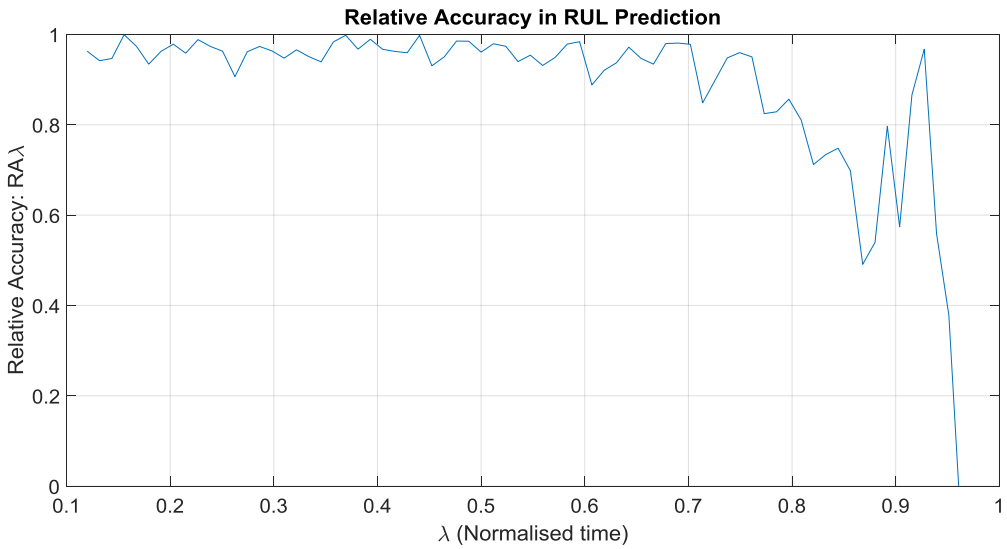





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900

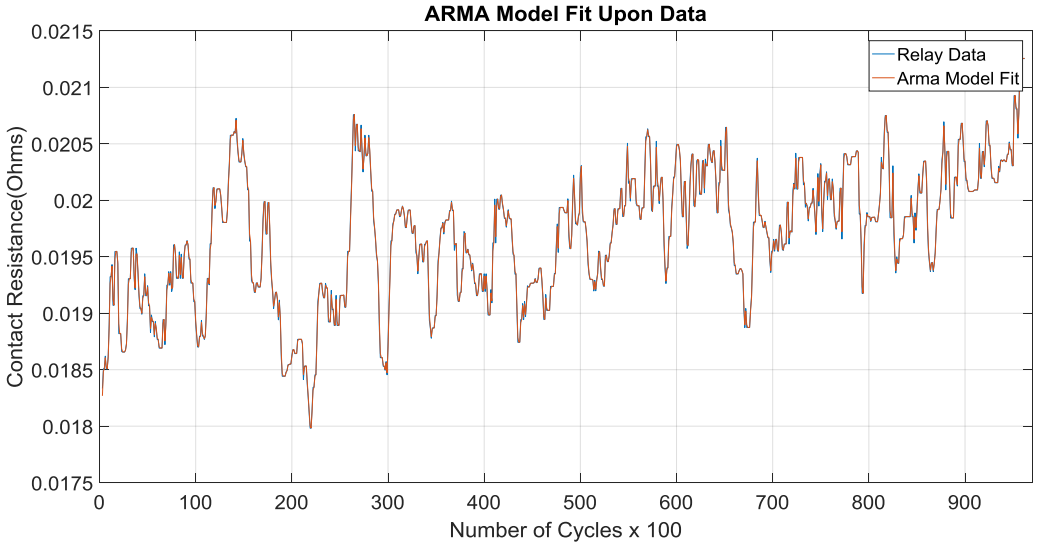

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
200
400
600
800
1000
1200

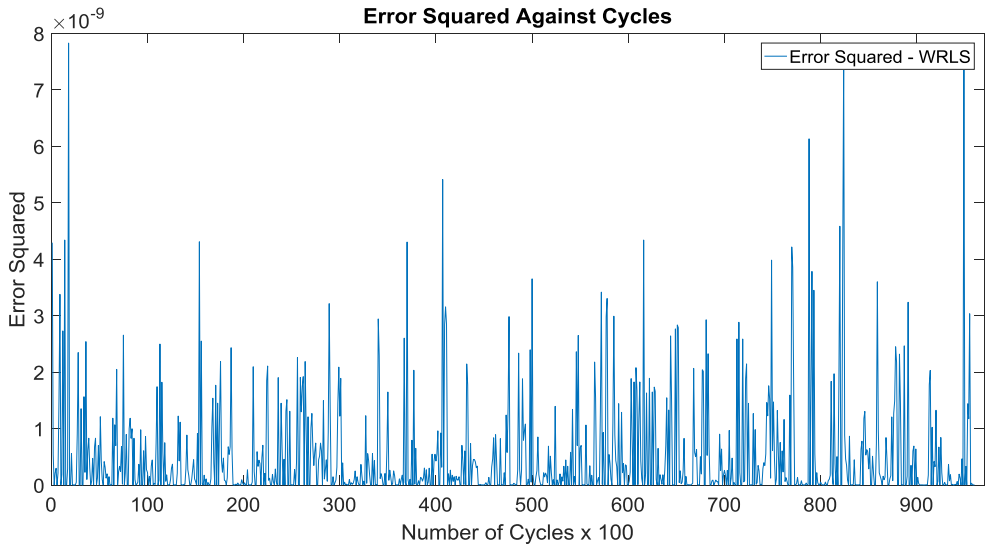

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
200
400
600
800
1000

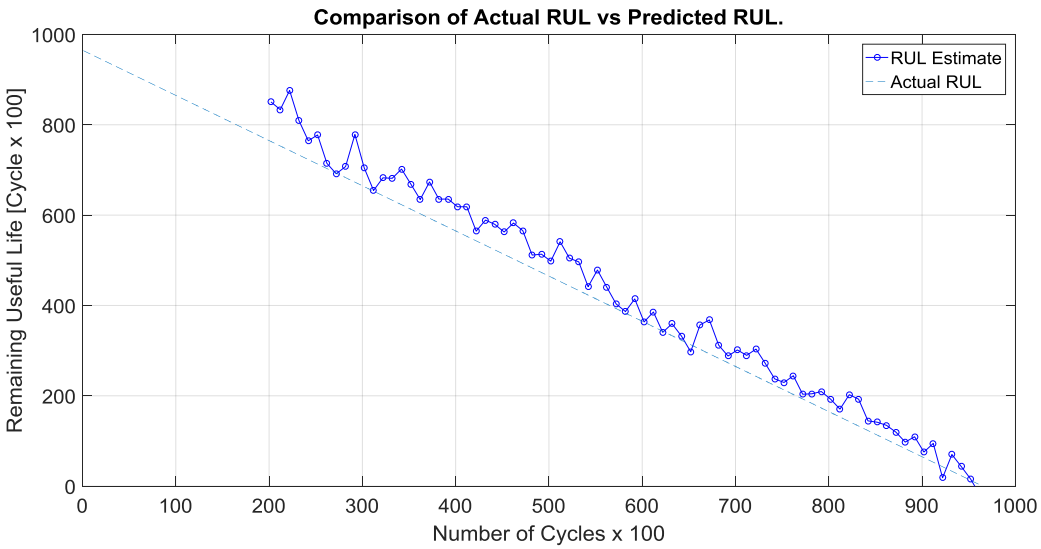


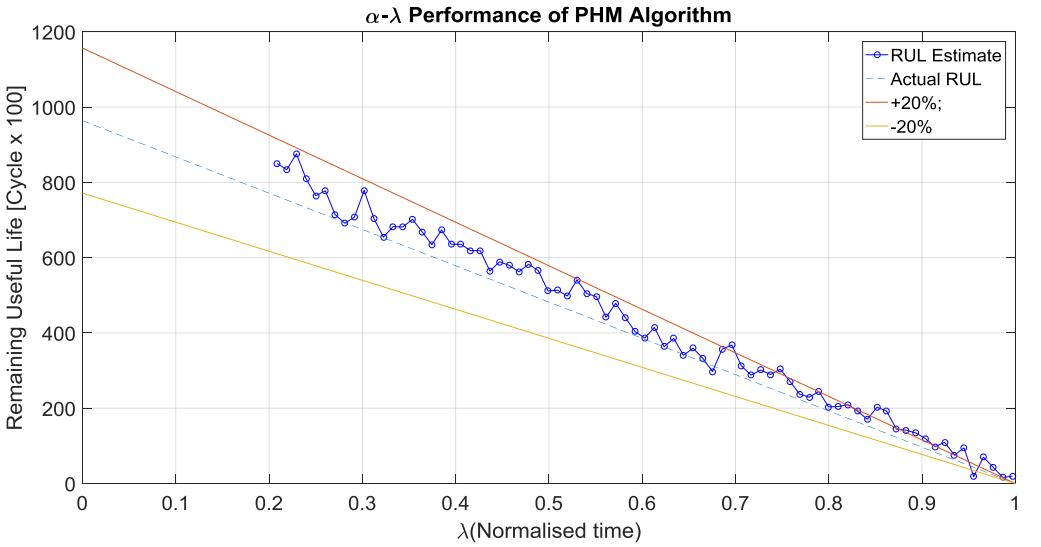
Figures 6.9. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 2 with 100 window.

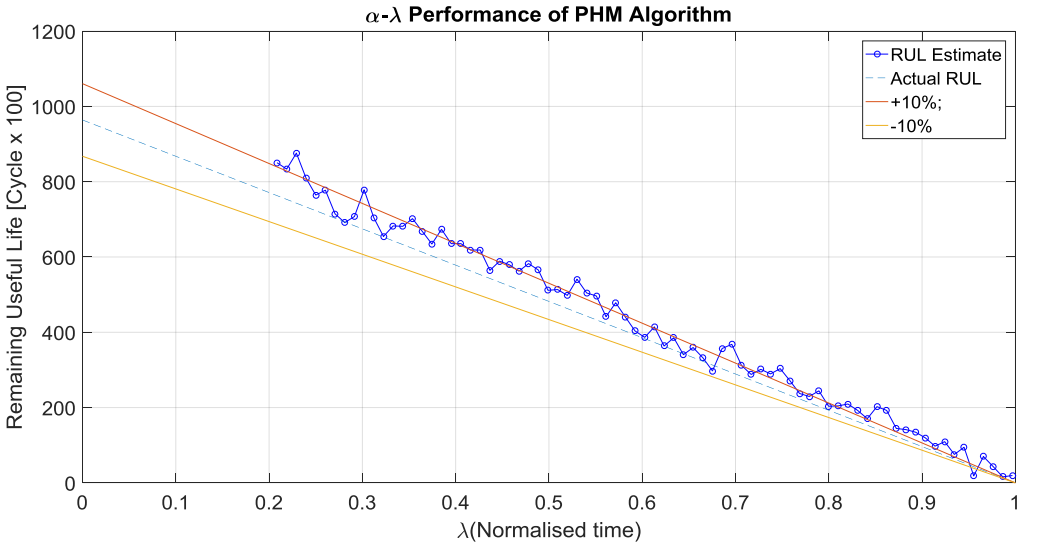
Sample 3 – 200

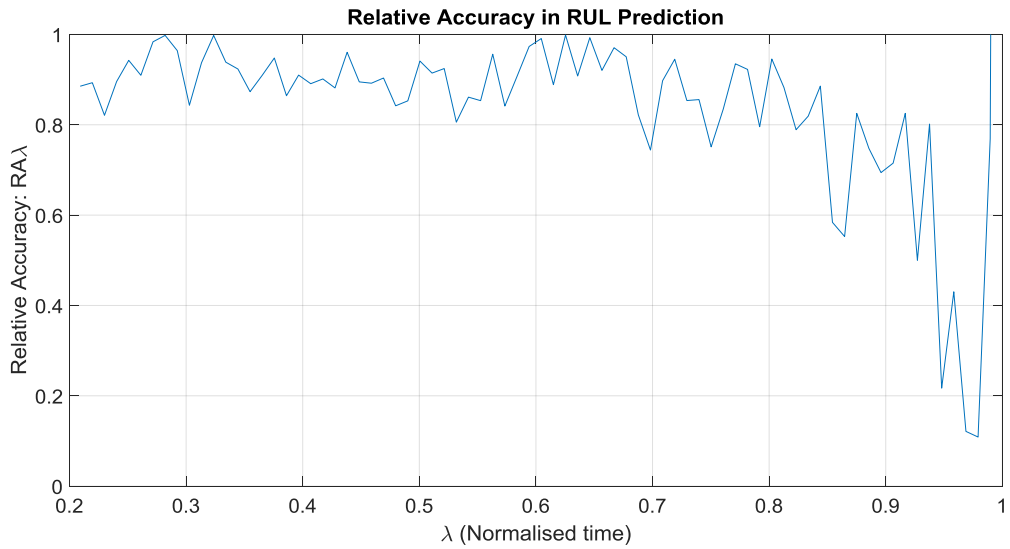





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000

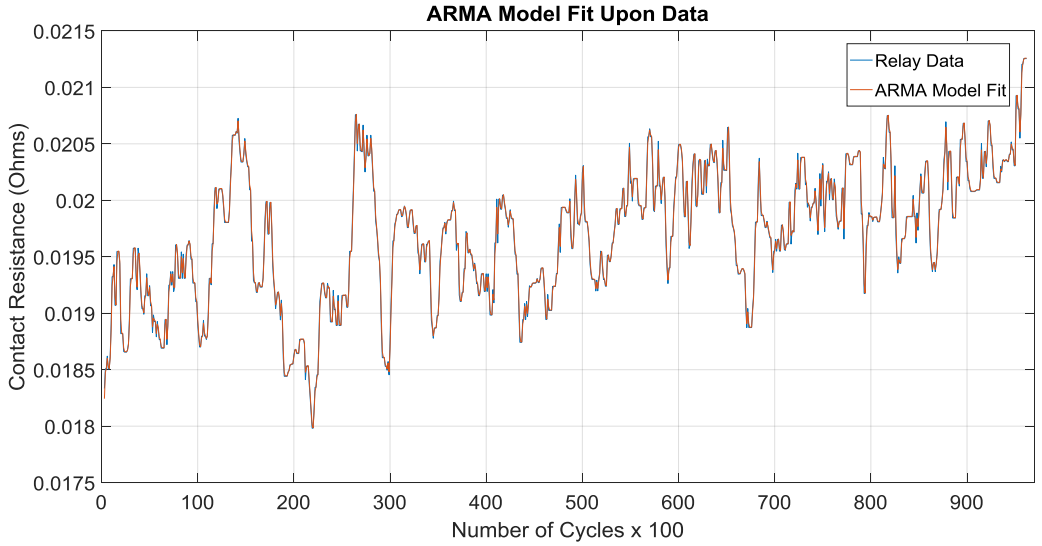

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

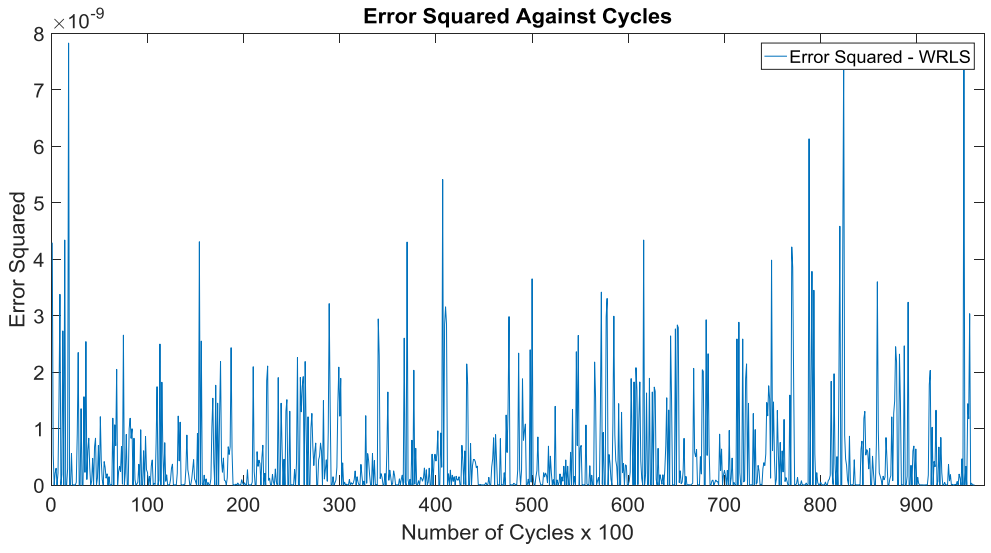

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

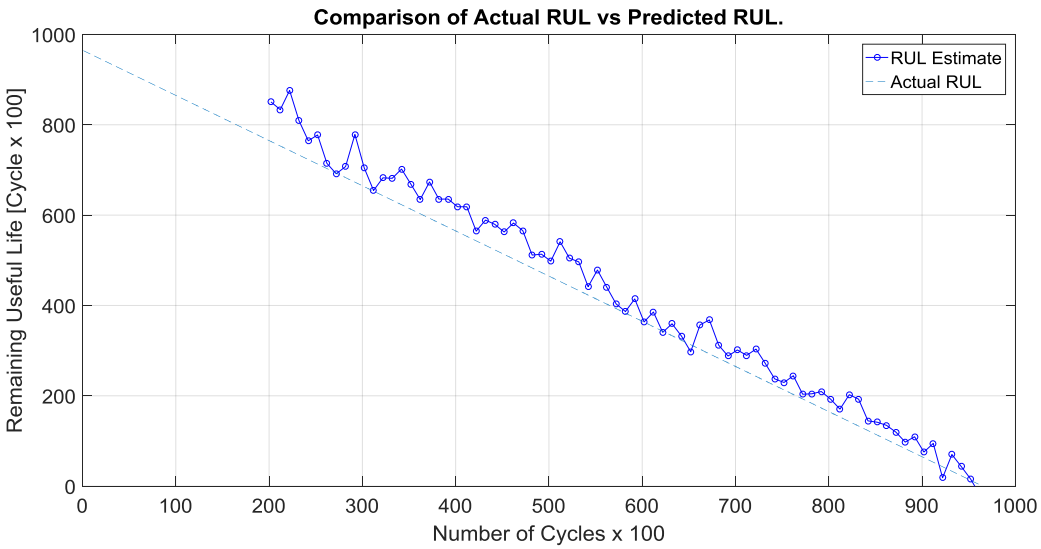


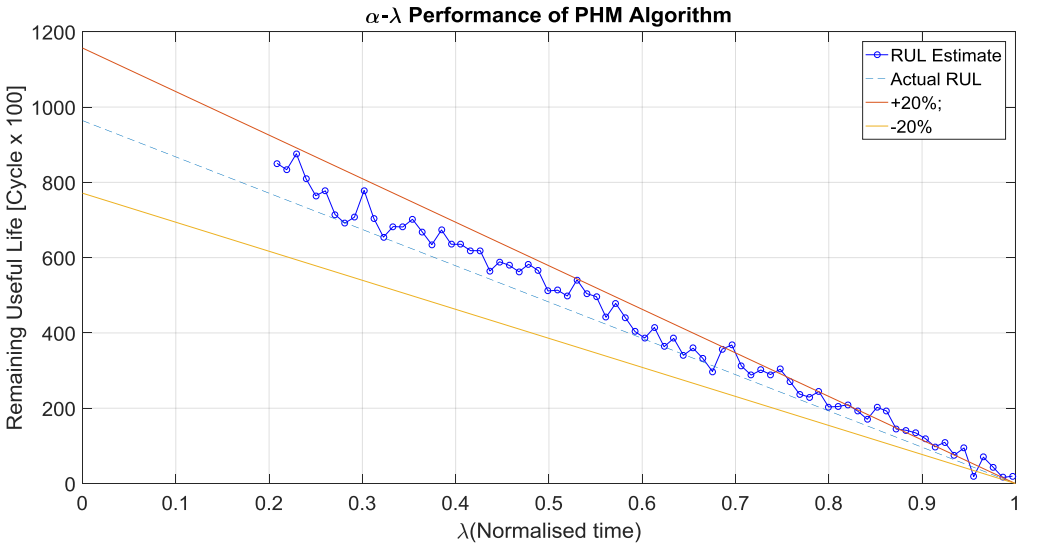
Figures 6.10. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 3 with 200 window.

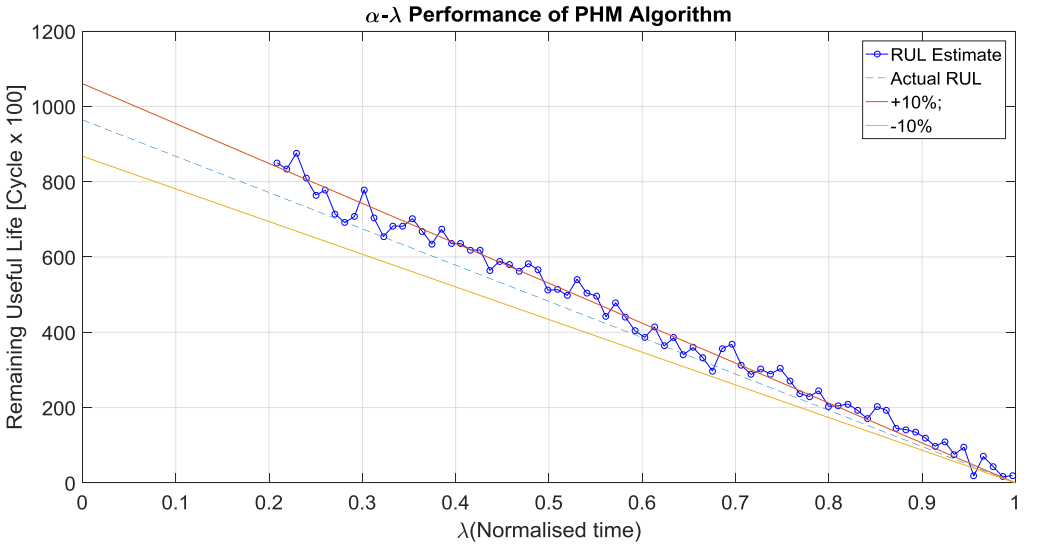
Sample 3 -100

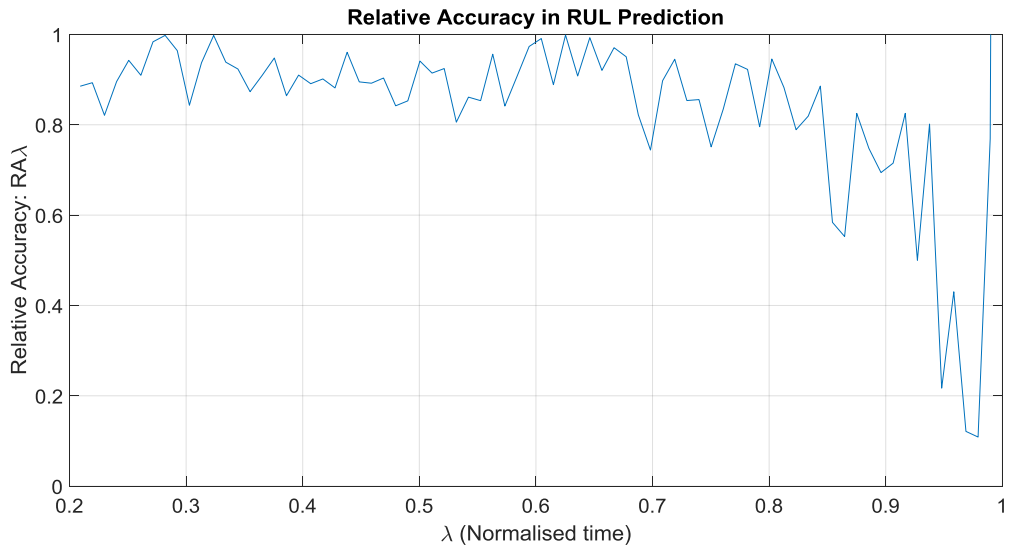





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000

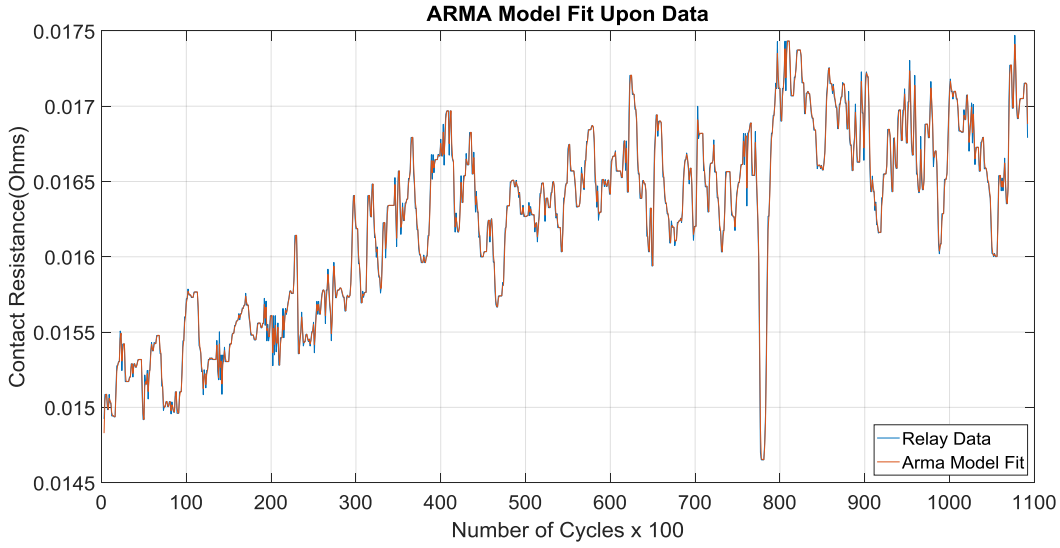

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

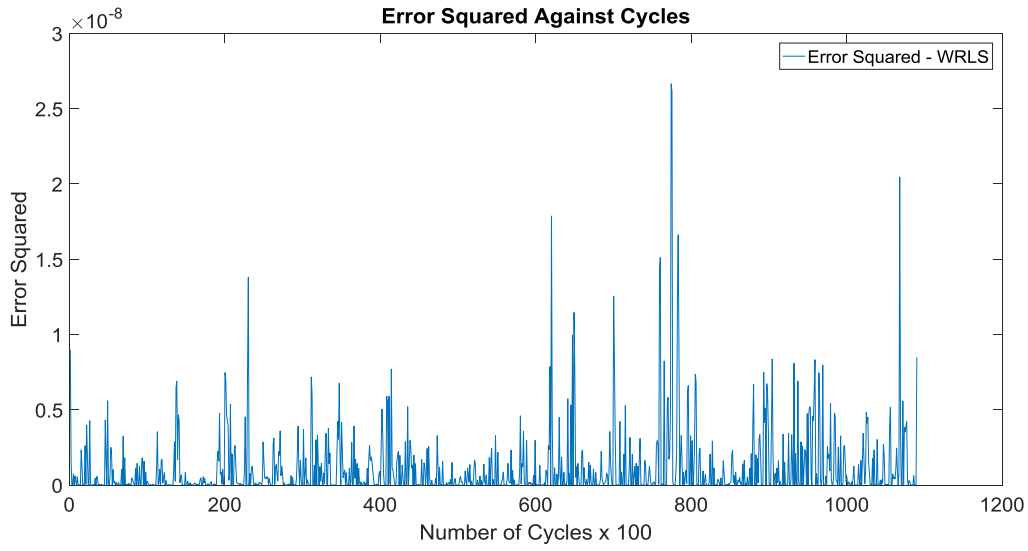

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

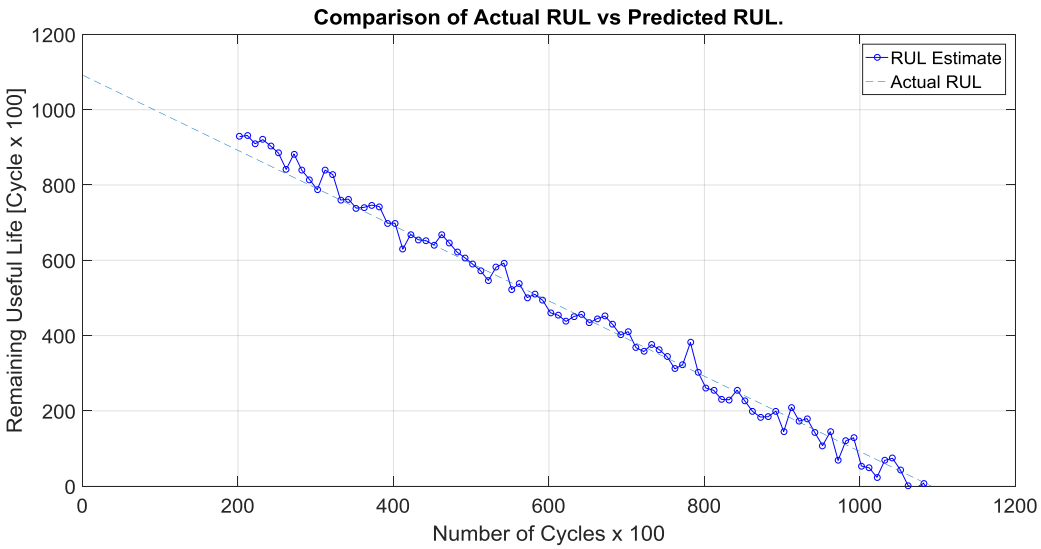


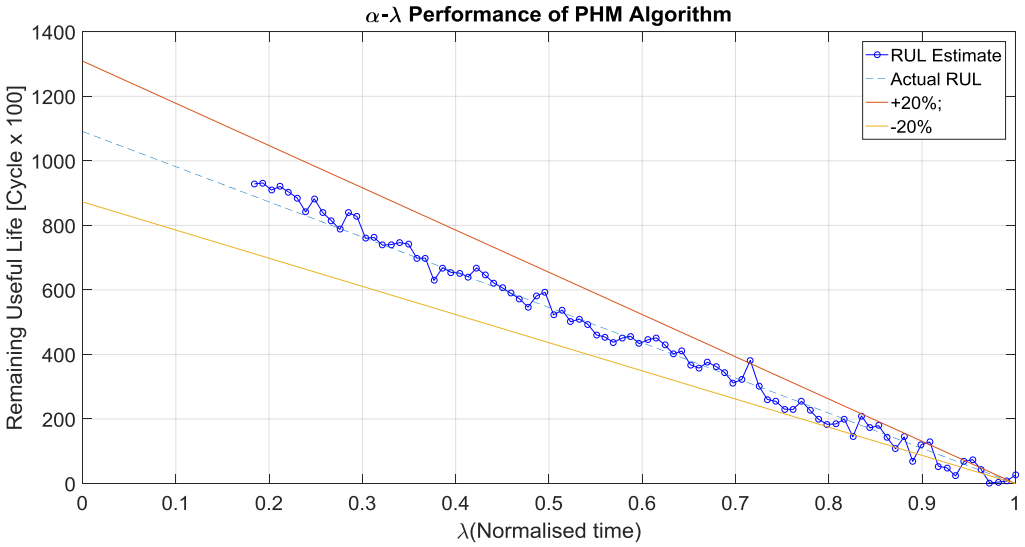
Figures 6.11. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 3 with 100 window.

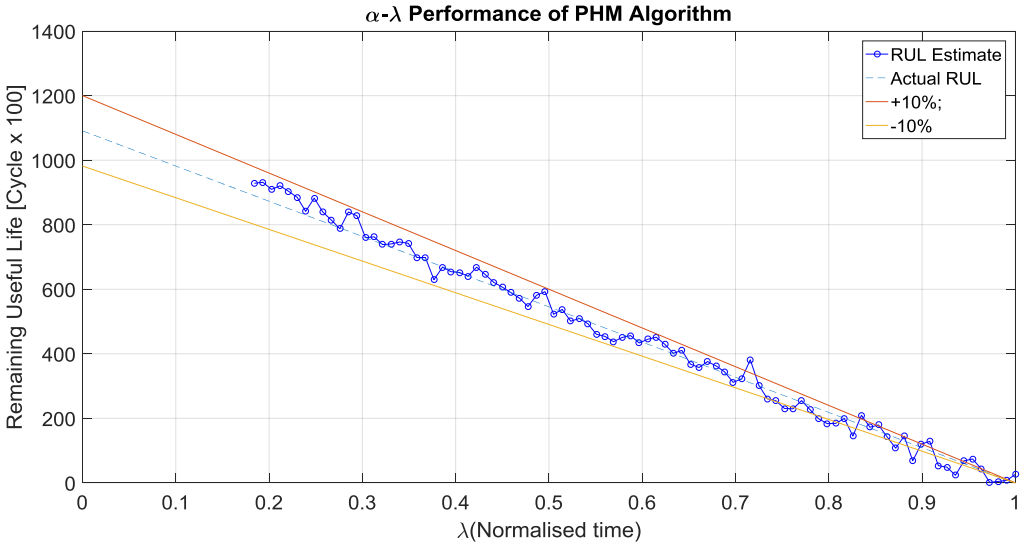
Sample 4 - 200

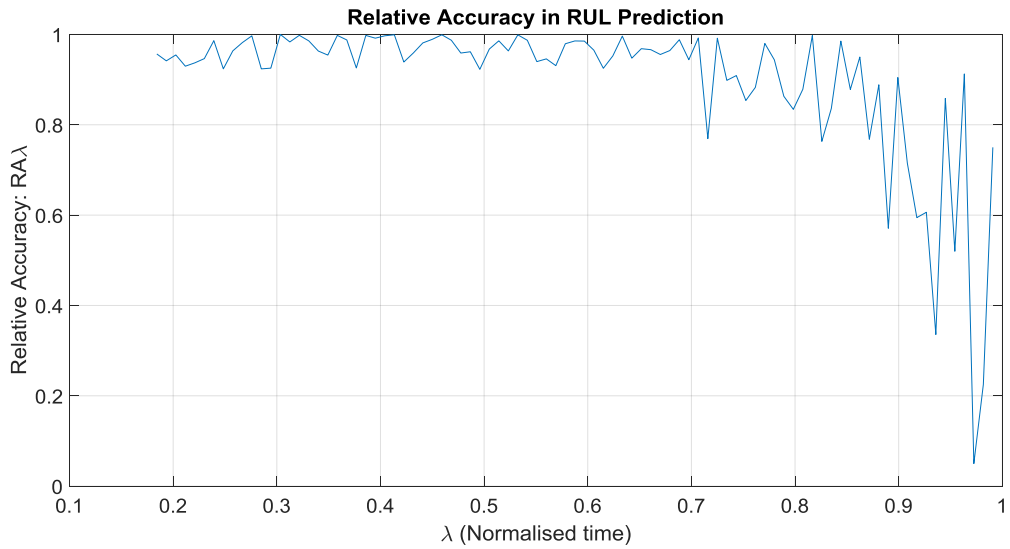





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
200
400
600
800
1000
1200

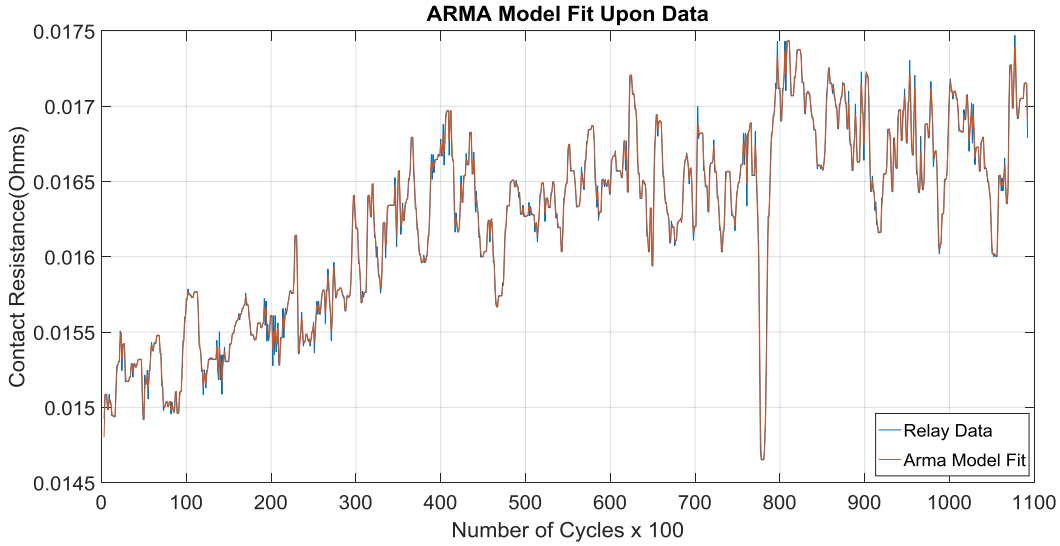

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
200
400
600
800
1000
1200
1400

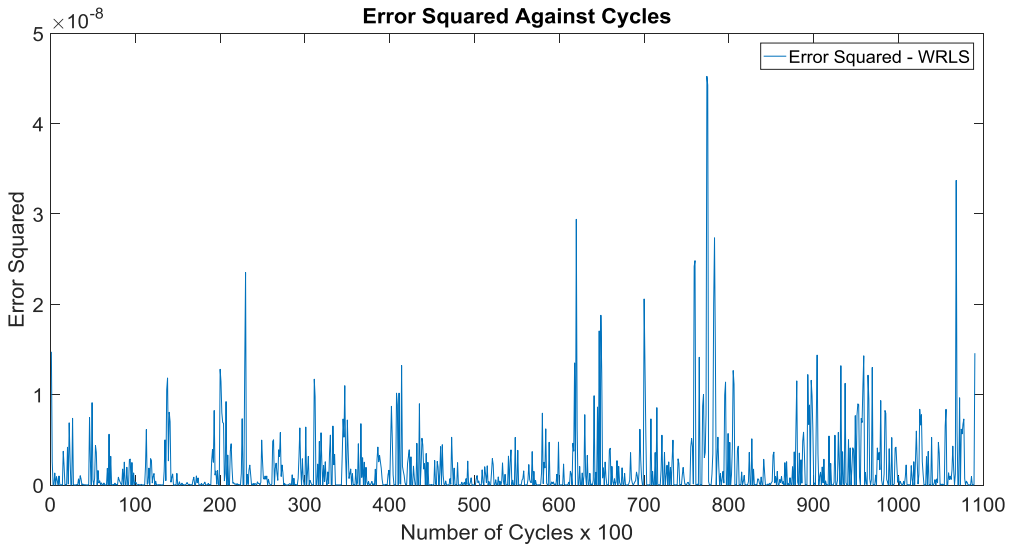

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
200
400
600
800
1000
1200
1400

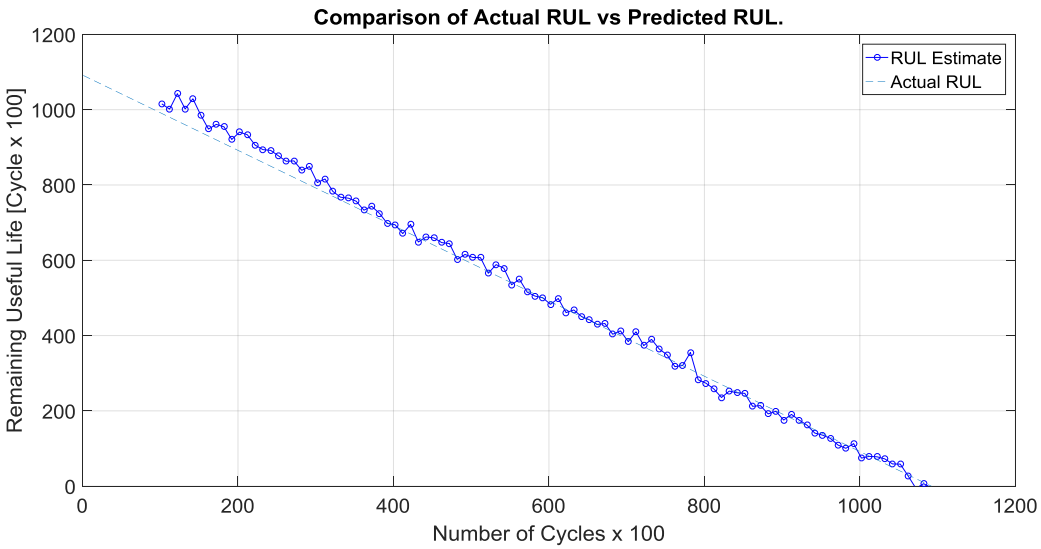


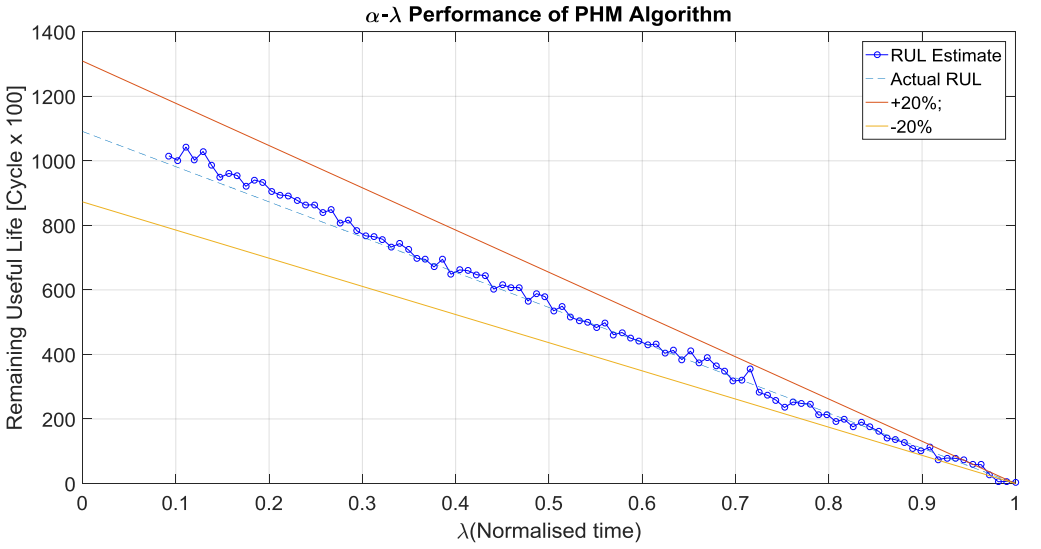
Figures 6.12. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 4 with 200 window.

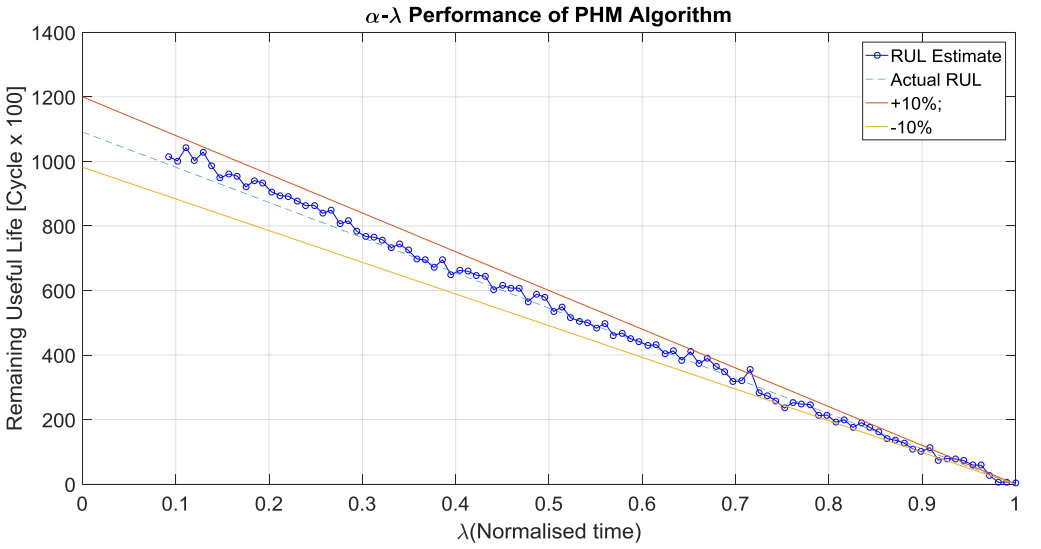
Sample 4 -100

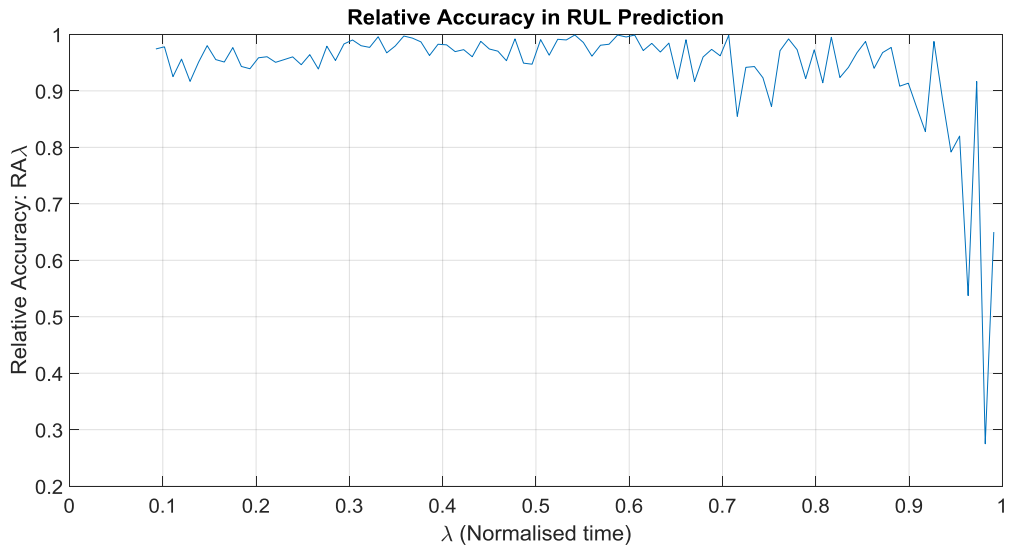





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
200
400
600
800
1000
1200

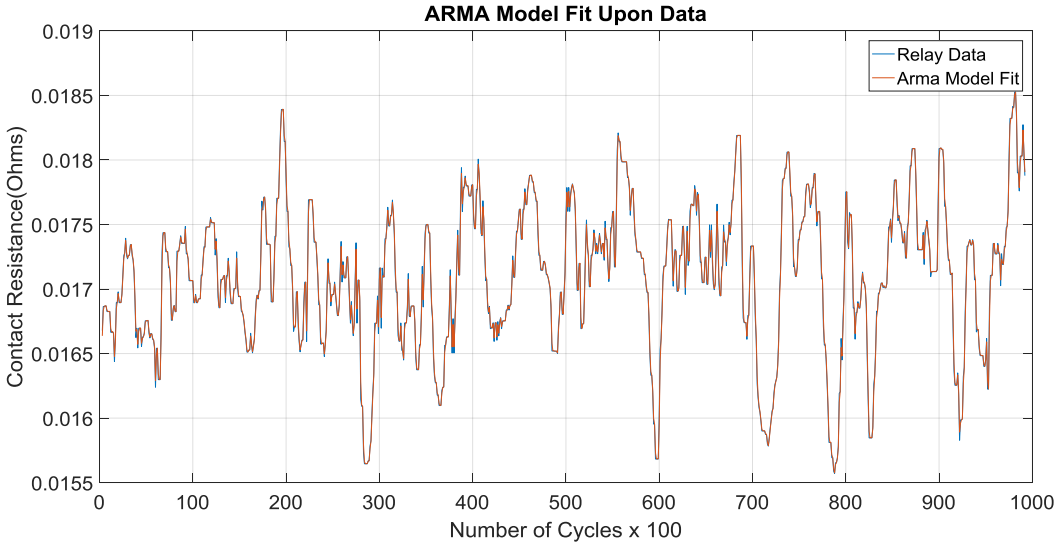

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1400

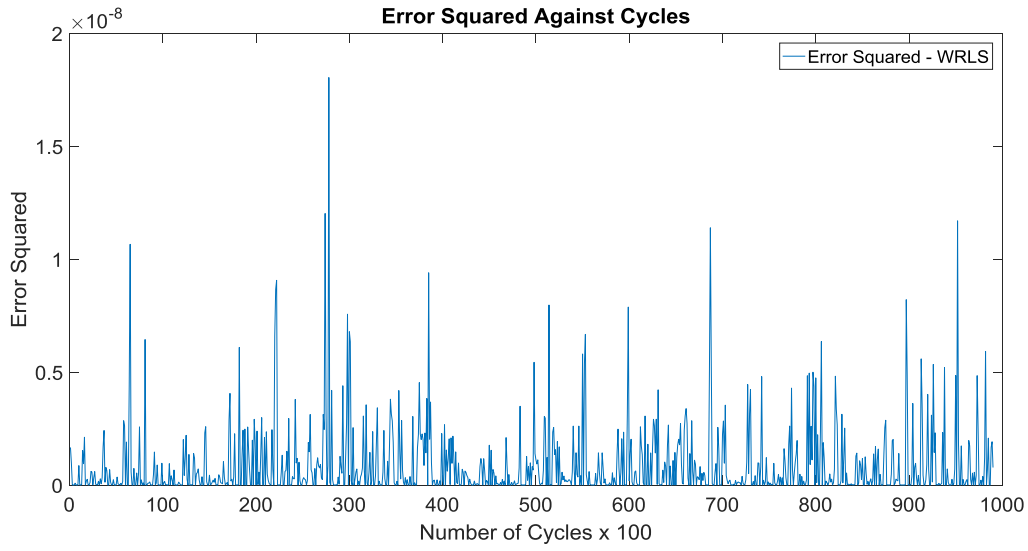

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1400

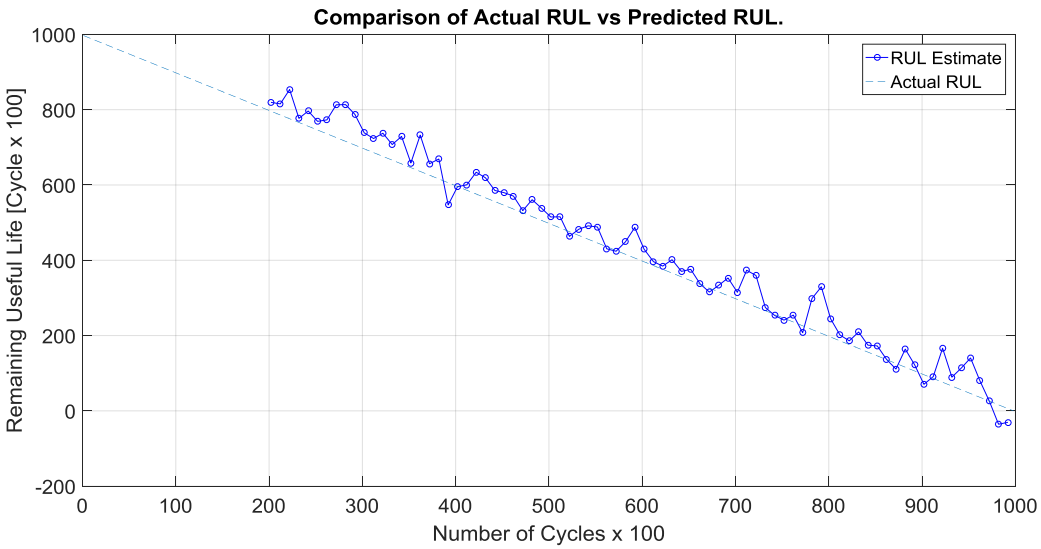


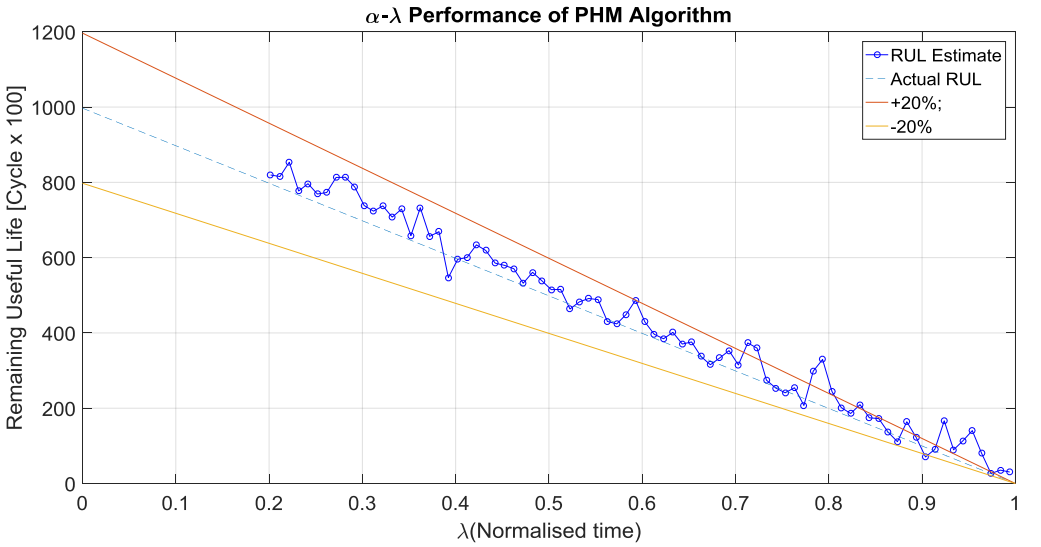
Figures 6.13. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 4 with 100 window.

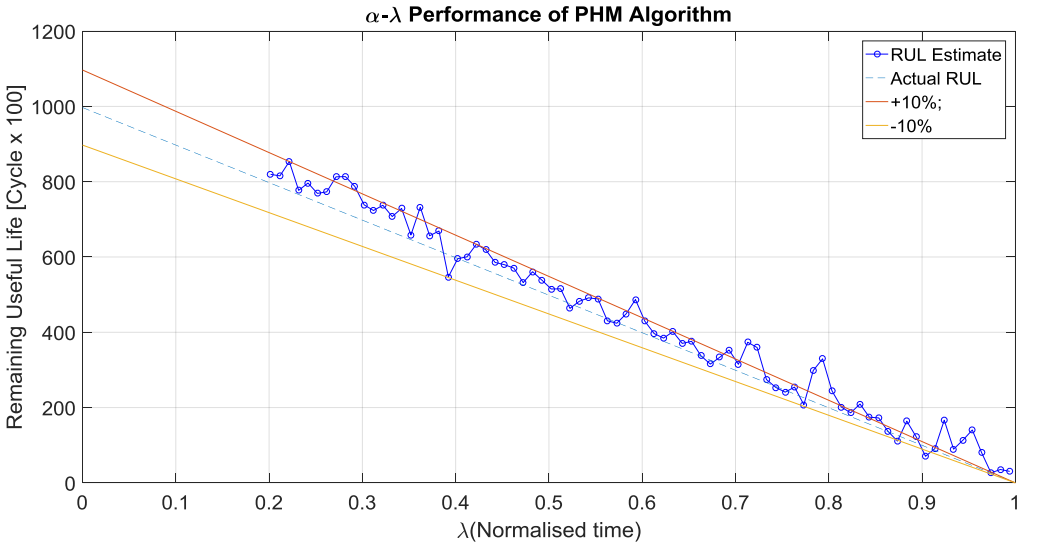
Sample 5– 200

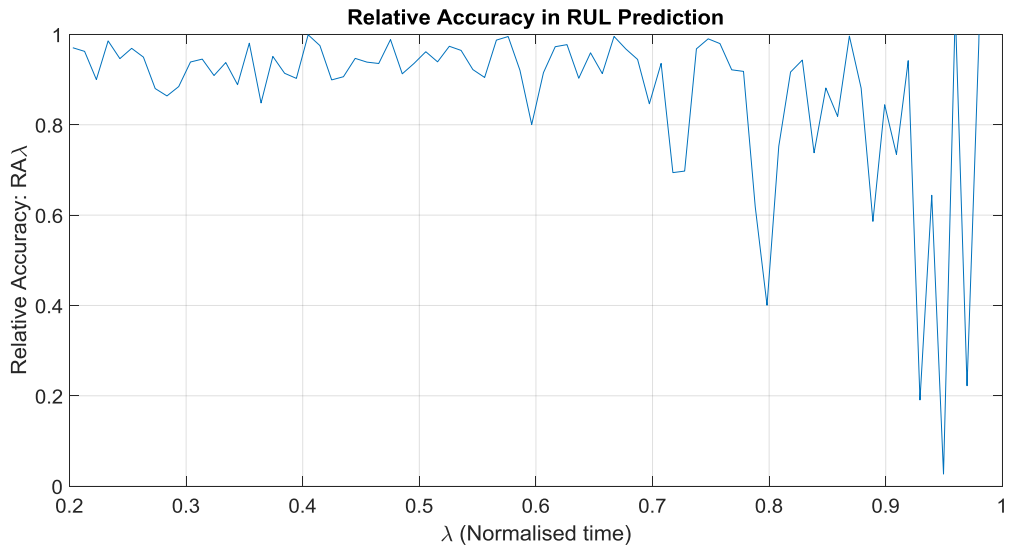





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
1000
800
600
400
200
0
-200
0
100
200
300
400
500
600
700
800
900
1000

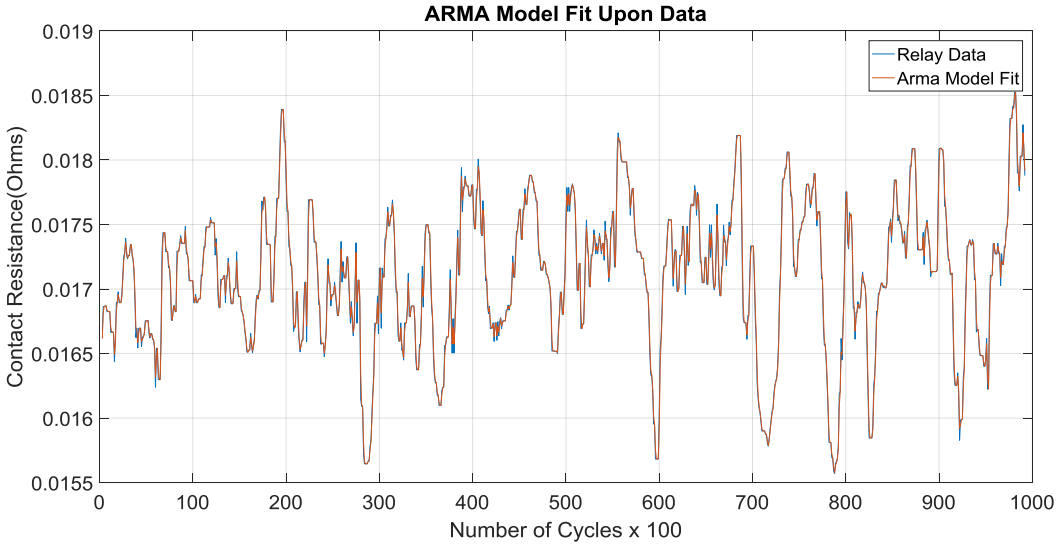

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
1200
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

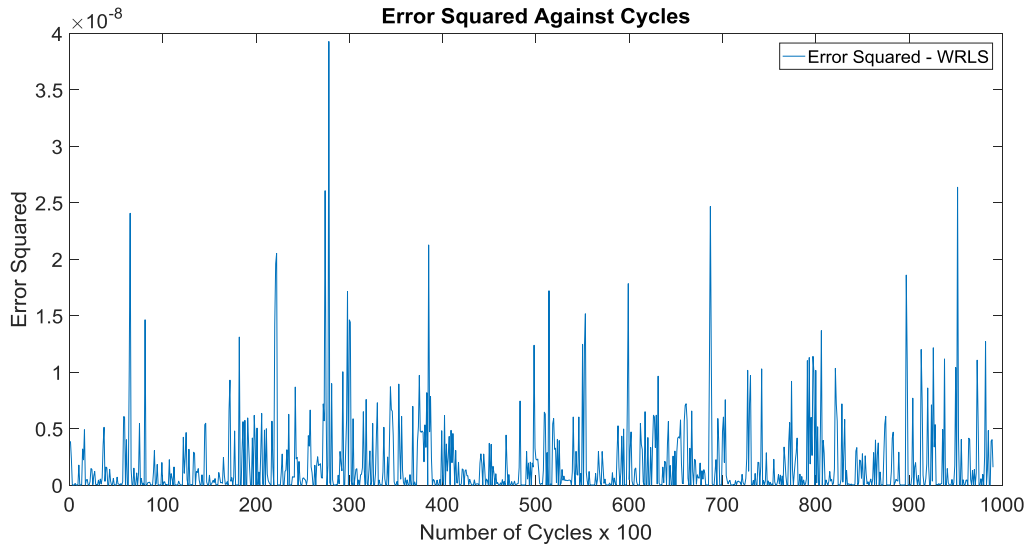

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
1200
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

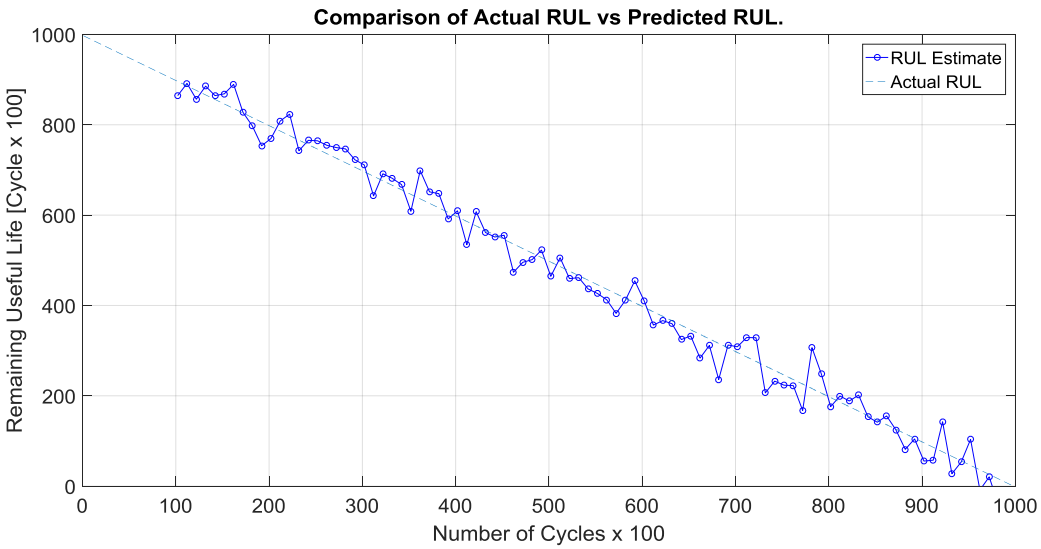


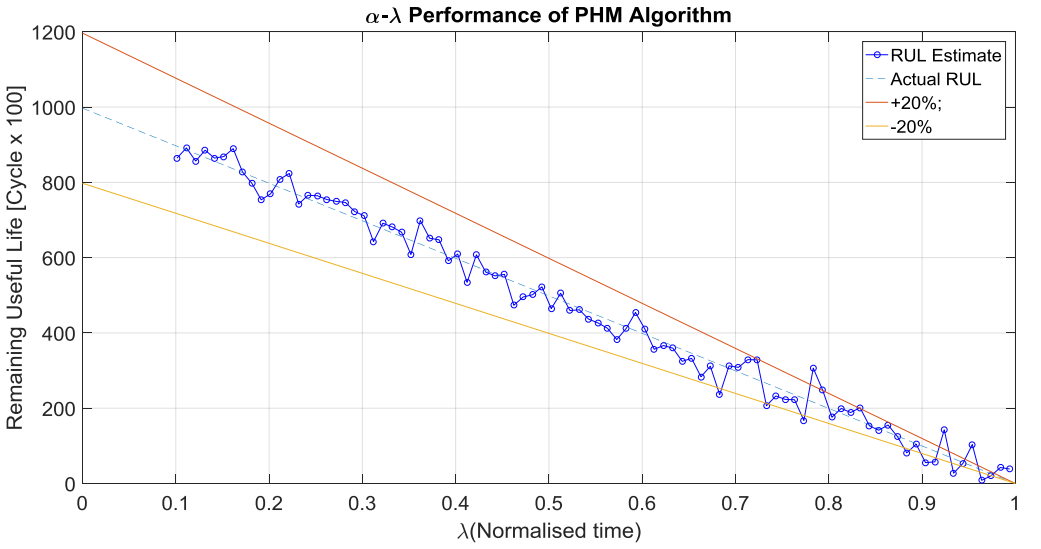
Figures 6.14. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 5 with 200 window.

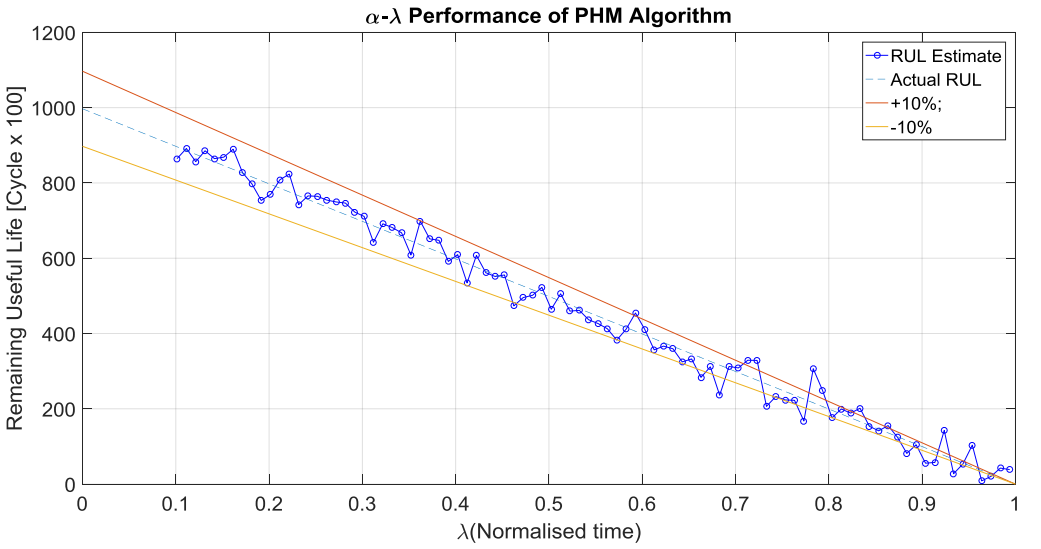
Sample 5 – 100

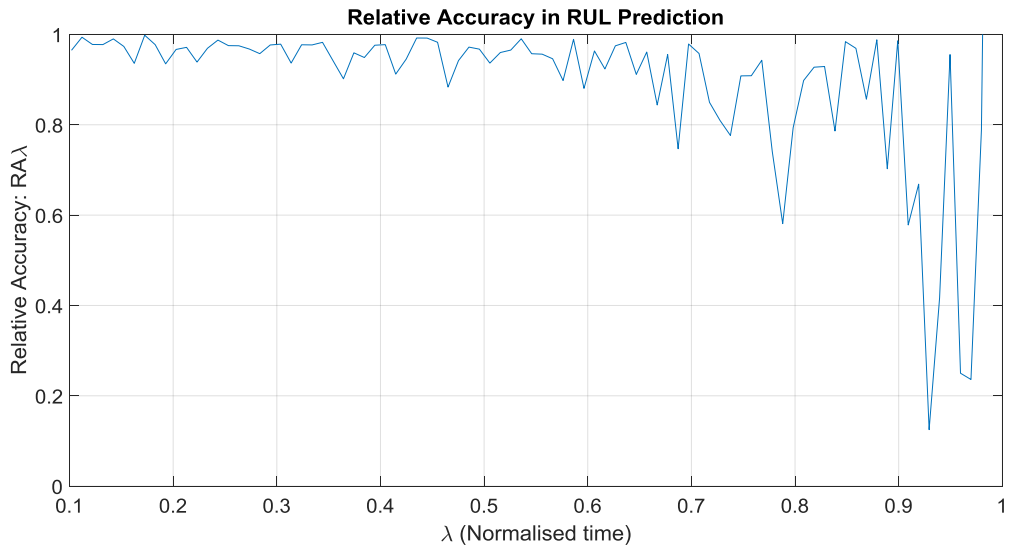





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000

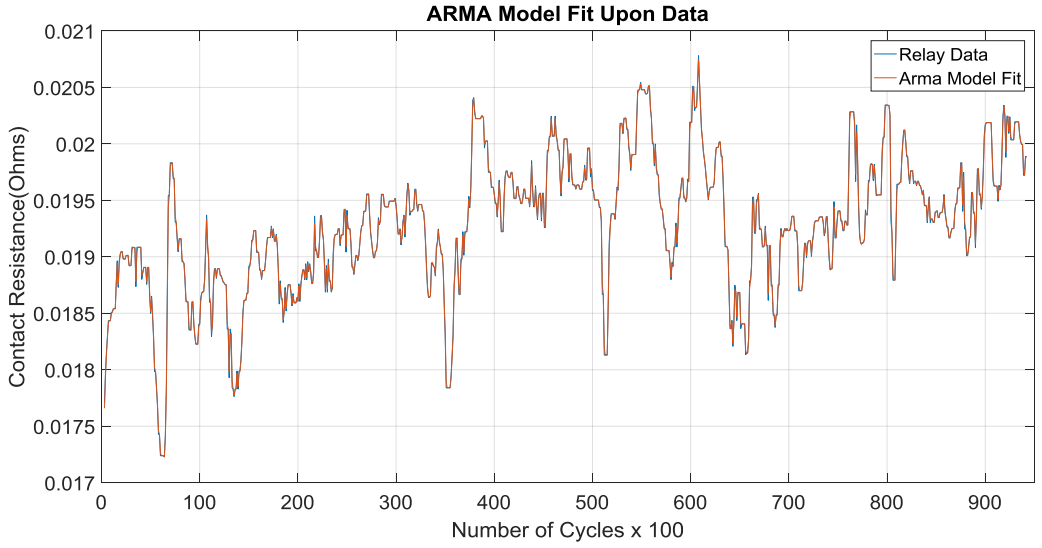

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

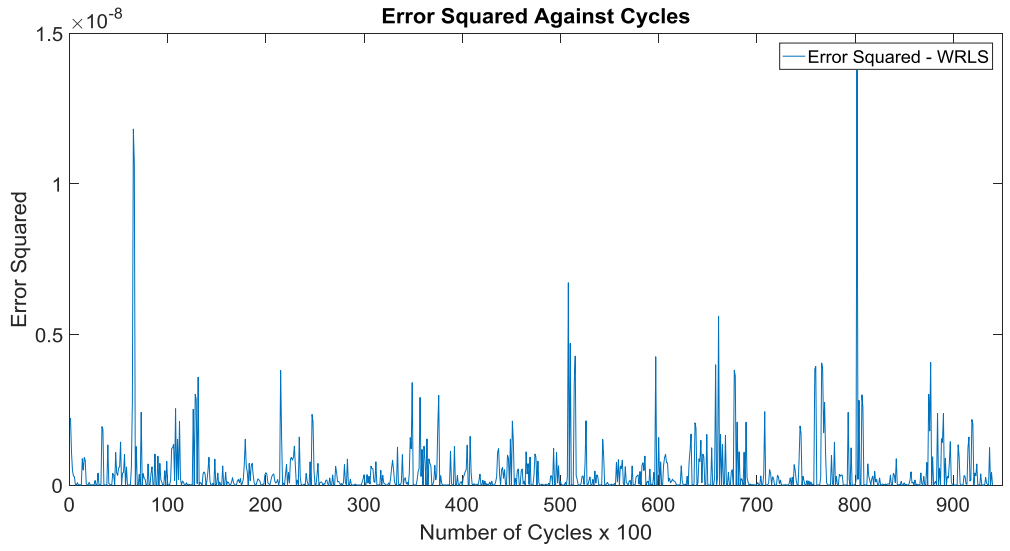

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1200

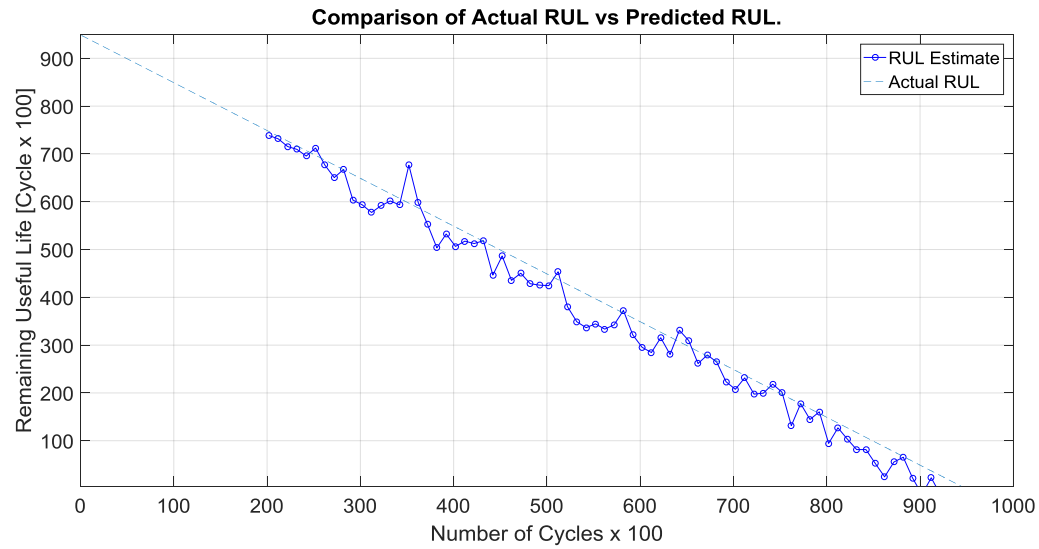


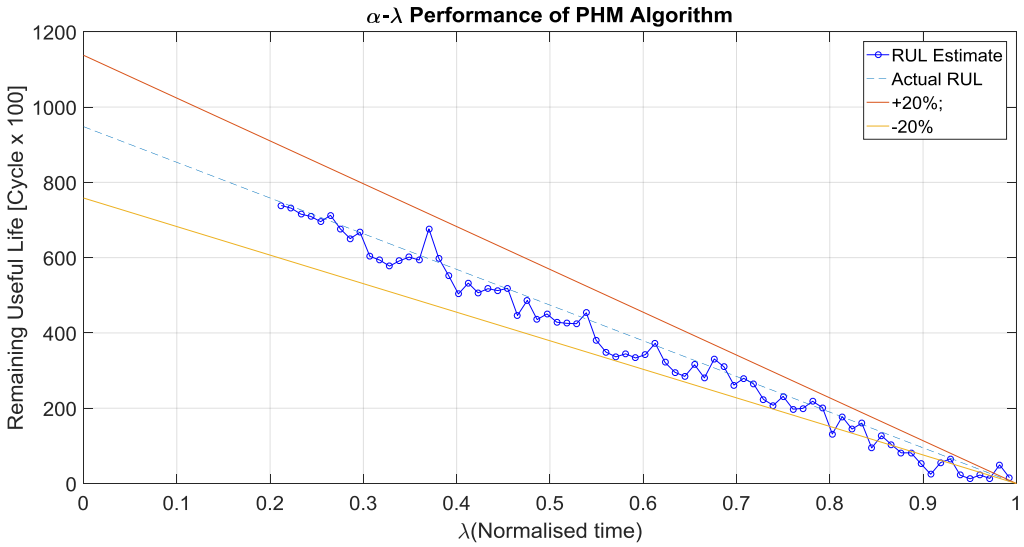
Figures 6.15. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 5 with 100 window.

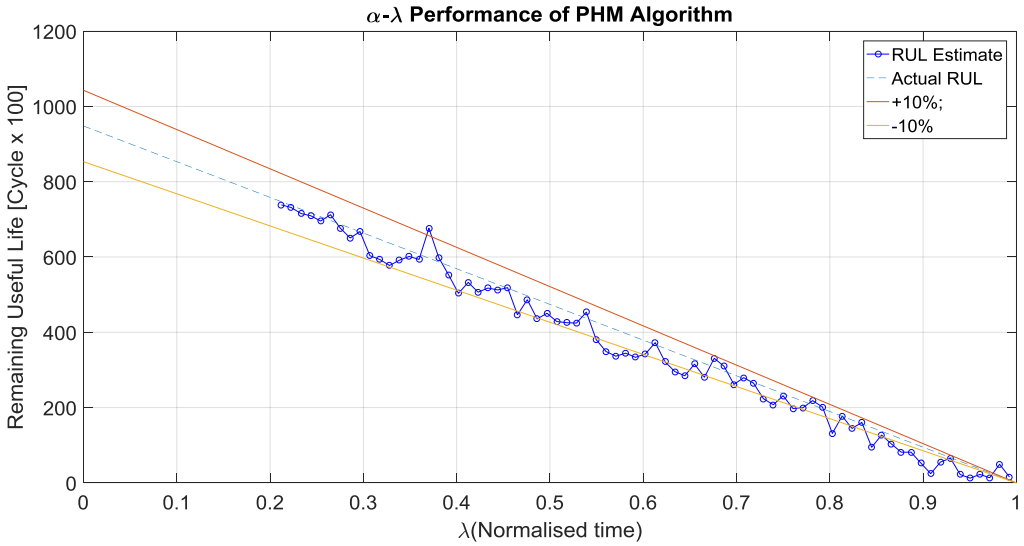
Sample 6 – 200

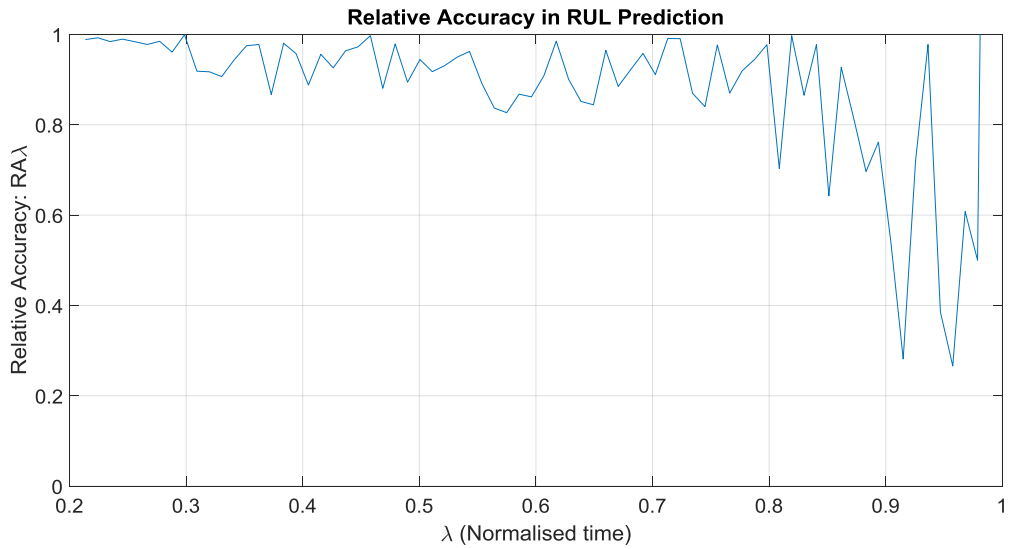





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100

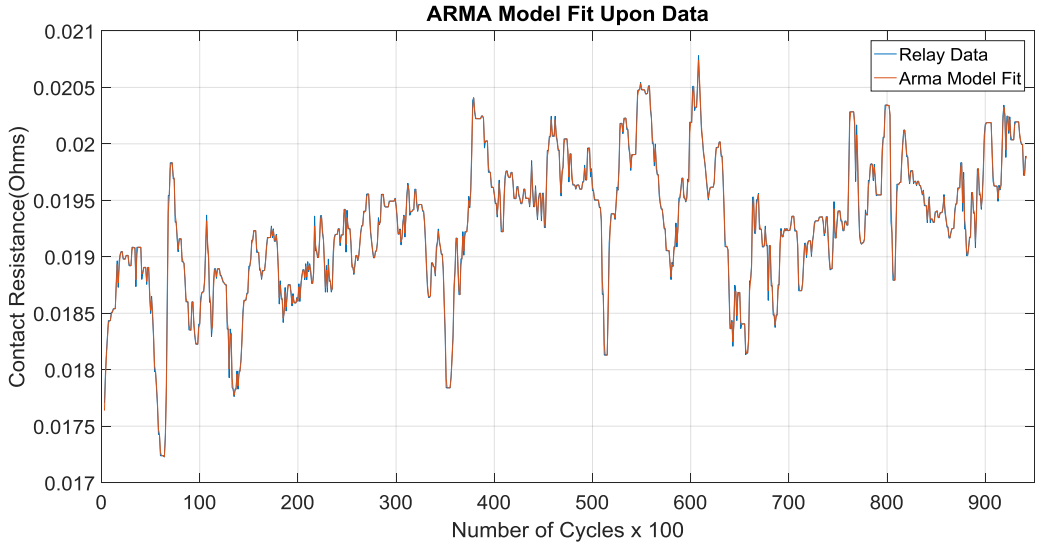

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)

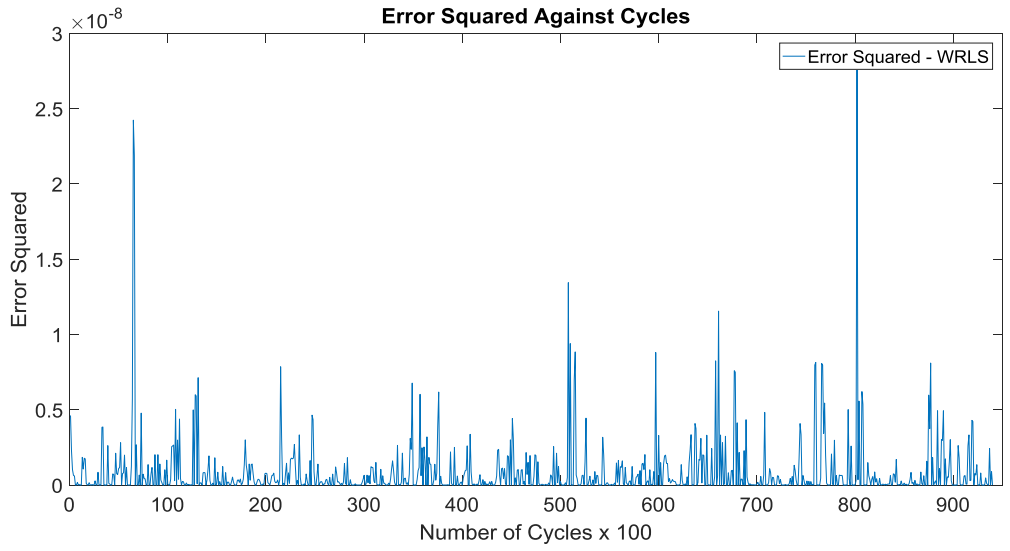

α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)

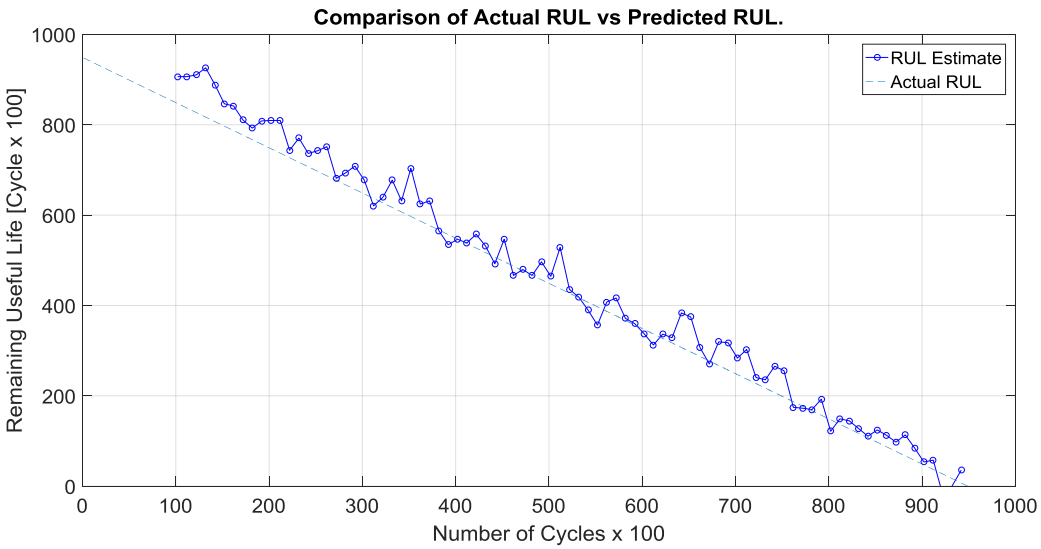


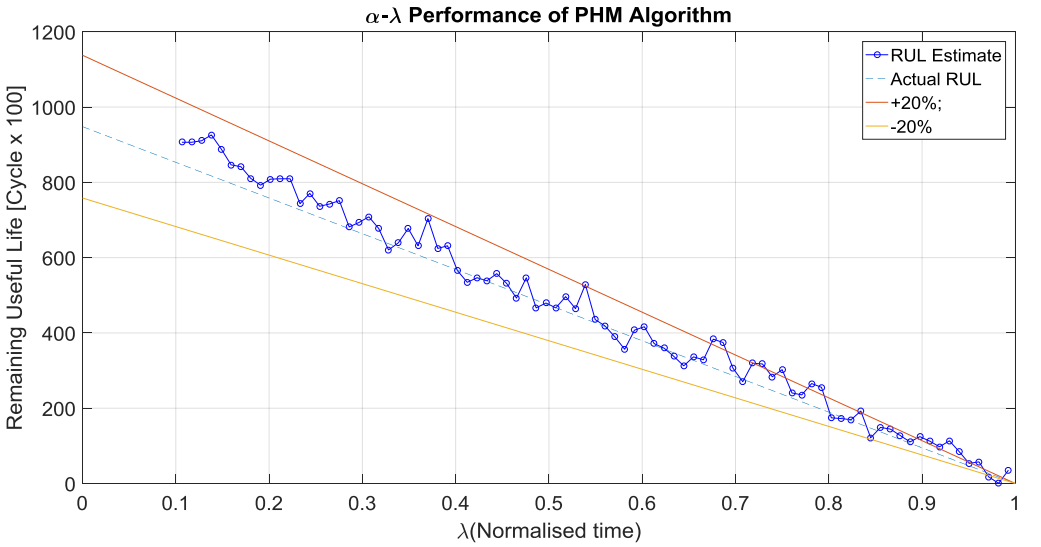
Figures 6.16. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 6 with 200 window.

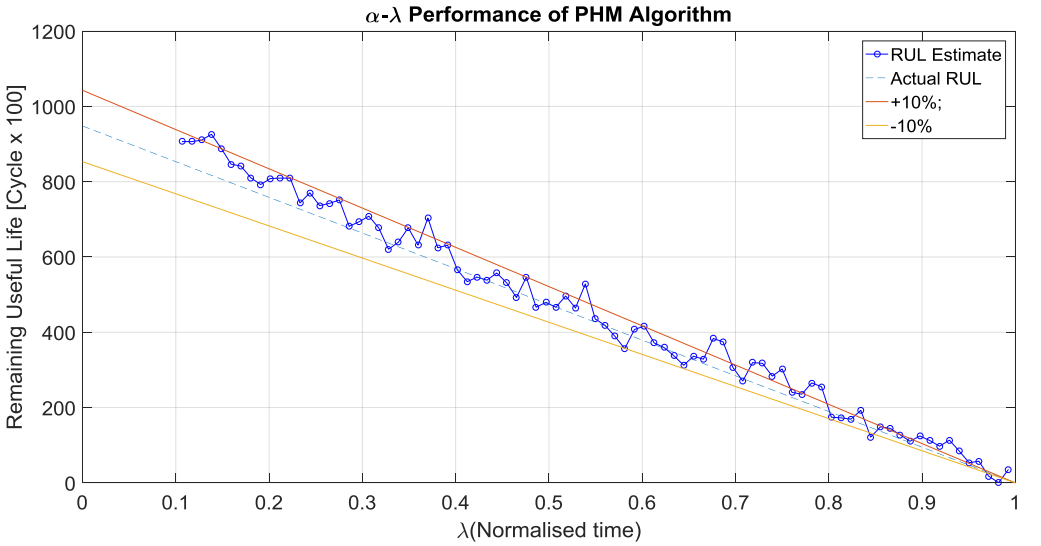
Sample 6 – 100

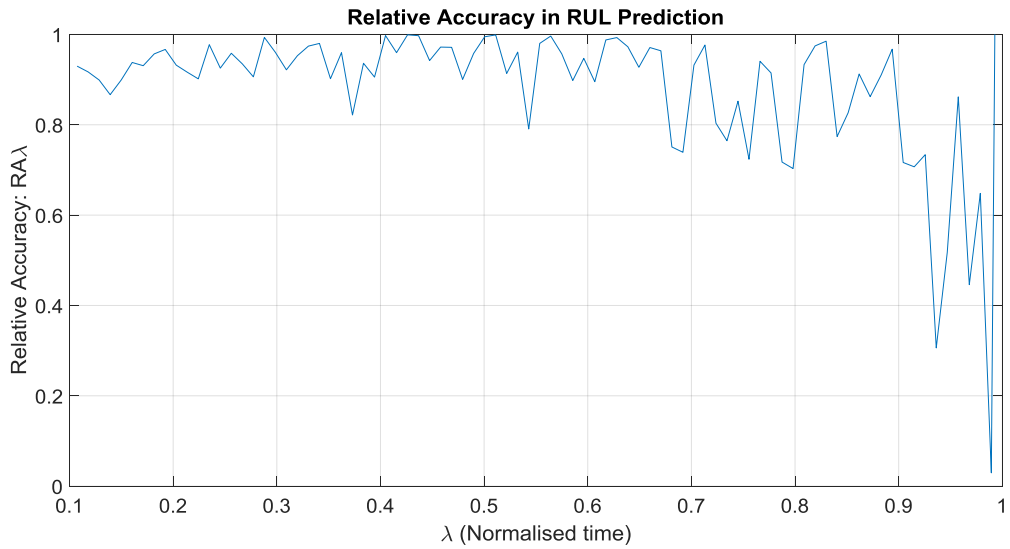





Comparison of Actual RUL vs Predicted RUL.
RUL Estimate
Actual RUL
Remaining Useful Life [Cycle x 100]
Number of Cycles x 100
0
100
200
300
400
500
600
700
800
900
1000
0
200
400
600
800
1000


α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+20%;
-20%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
200
400
600
800
1000
1200


α-λ Performance of PHM Algorithm
RUL Estimate
Actual RUL
+10%;
-10%
Remaining Useful Life [Cycle x 100]
λ(Normalised time)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
200
400
600
800
1000
1200



Figures 6.17. Showing ARMA fit, Error, RUL, $\alpha - \lambda$ at 20%, $\alpha - \lambda$ at 10% and RA for sample 6 with 100 window.

Conclusion

A computationally efficient, technique has been developed for real time monitoring of accrued damage in electrical contacts within safety critical applications. The technique utilizes Recursive Least Squares parameter identification via a sliding window and a Bayesian PHM framework in the form of a Kalman-filter for state variable evolution. The estimated state-space parameters were used to extrapolate the feature vector into the future, and predict the time-to-failure at which the feature vector will cross the failure threshold. This procedure was repeated recursively until the component failed.

Remaining useful life was calculated based on the evolution of the state space feature vector. Standard prognostic health management metrics are used in chapter seven to quantify the performance of the algorithm against the actual remaining useful life.

This chapter provides the following conclusions and contributions

- The development and assessment of a real time Grey Model prediction algorithm for prognostics.
- A novel and computationally efficient technique has been developed for real time monitoring based upon modelling of degradation using system identification techniques.
- Recursive Least Square (sliding window) for state variable extraction
- Bayesian Methods (Kalman filter) for prediction of RUL

Chapter 7 - Results from PoF and SWRLS ARMA Models

Introduction

This chapter presents the results from the models derived in chapters 5 and 6 respectively. The Physics Based Model is firstly analysed and discussed and then compared to the Data Driven model. In order to make the comparison between the two models, a form of metric is required to enable benchmarking to be carried out. This chapter starts by outlining current prognostic metrics and the results are evaluated according to these metrics. The chapter concludes by discussing the outcomes of the results.

7.1 Physics Based Modelling

The detailed discussion of the methodology and derivation of the physics of failure based model is documented in Chapter 5. This chapter outlines the results of the Remaining Useful Life (RUL) assessment based on this model, including the use of the metrics discussed above in benchmarking the prognostic performance of the model. Six sample data sets were used in the proving the model, the data was obtained from the experimental methodology discussed in Chapter 4.

7.1.1. Model

The state model representing the degradation of the relay contact resistance over time is combined with a Particle filter to project the degradation into the future in order to form an estimate of the RUL. To help visualise the process, the output from the Particle filter illustrated in figure 7.1, along with the histogram representation of RUL and the actual data superimposed in the background. The current measurement value along with auxiliary parameters such as measurement error are tracked as part of the learning process and updated. Starting from the RUL estimation point, where the measurement input ends, the model parameters are extrapolated into the future up to the maximum contact resistance threshold using the discretised contact resistance state equation via Monte Carlo simulation.

From this projection, the RUL distribution at the specified threshold point is obtained by calculating the difference between the time of the last measurement (RUL estimation start point time) and the time where the projected trajectory hits the

threshold, this distribution will be equal to the number of particles at that threshold. The blue line in Figure 7.1 represents the mean value of the distribution, where the red boundaries represent a confidence encapsulation of 95% of the distribution, forming a confidence boundary.

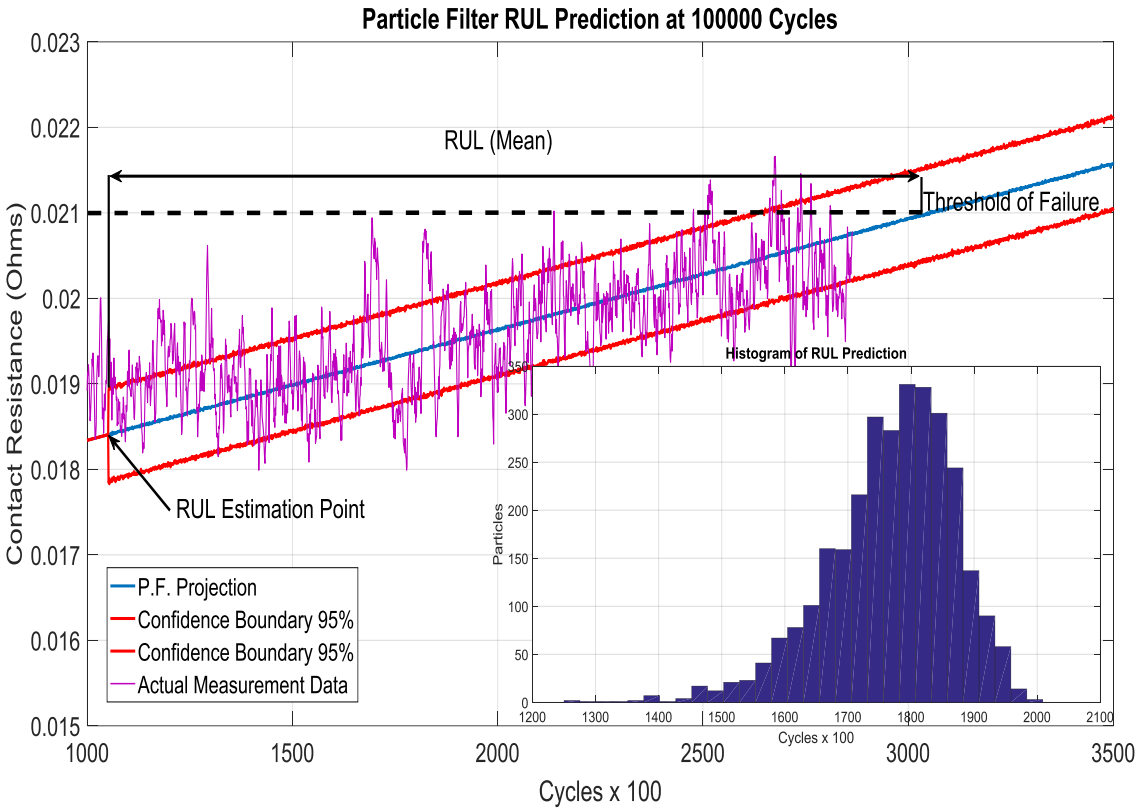


Figure 7.1 The Particle filter projection against the actual measured data and histogram representation of RUL.

For each sample, the RUL estimation was performed at every 5000 cycles. The RUL is shown for each sample below along with the alpha-lambda and relative accuracy projections. To help the visualisation of the models RUL and alpha-lambda metrics a shrinking alpha cone visual comparison, recommended by (Saxena et al., 2010) is used. The results from the model should stay within the 20% error bound predominantly throughout the degradation process.

The standard deviation value for the measurements 'σv' in the Particle filter was 0.003 and in each sample, the number of particles employed in the filtering was three thousand respectively.

The results for samples one to six are shown below with their respective RUL, alpha-lambda and relative accuracy plots as well as the Particle filter prediction and histograms of RUL for samples 1, 4 and 6 to help visualize the prediction process.

7.1.2 Physic of Failure Model Discussion

The output from the model is linear in nature, and although it is not an exact fit to the data, the estimation is, within most cases 20% accurate, if not better. This becomes apparent if a best fit line is drawn though the data sets up to the threshold of failure, if one neglects the deviations in contact resistance due to the film formation and rupture and material transfer as discussed in Chapter 4, the degradation data for the most part approximates to a first order polynomial. Figure 7.1 above, illustrates the Particle filters approximation to a set of data, with a histogram of RUL.

From the results in Chapter 5, samples 1 and 4, the Particle filter shows the prediction at various cycles and shows the projection up to the threshold of failure. Generally, the first 100,000 cycles will consist of a period where the contact resistance will either slightly decrease or remain steady for the first few tens of thousand cycles and then linearly increase; in sample 1 there is a rise up to 50,000 and then a subsequent drop. This outlines the problem of developing a physics model that encompasses all scenarios likely to be encountered when the failure mode is complex and consists of multiple degradation parameters. Usually, physics based models may be bolstered with empiric data gathered from experiment results, however even that is of little use here. So a projection up to the failure threshold using the predominant failure mode was deemed to be the best fit for the data.

Not being able to represent the model accurately at all stages will inevitability impact upon the performance of the prognostic. Samples 1 and 6 show the initial representation is quite good, well within the 20% tolerance band, this is due in part to the linear nature of the data up to this point and to ability of the particle filter to adapt very quickly to the data from the simplistic nature of the model. From the RUL and α-λ plots, the prediction starts to fall out around half way along, this correspond to where the measurement data fluctuates considerably and the linear model can no longer make an accurate estimate, which is also depicted in the RA graphs.

The results from samples 2, 3 and 4 are a great deal better due to the data not being so volatile and the model fitting with a great deal more confidence. This is reflected in the metrics, as most of the data is within the 20% cones in both the RUL and α-λ plots throughout, the RA is considerably higher throughout.

The conclusion may be drawn that to represent the degradation accurately, the data needs to be a relatively good linear approximation, and the small anomalies due to variation in measurement will be handled by the Particle filter. However, where large variations outside the linearity of the model take place, accuracy is affected considerably, which would be a cause for concern in general use. The possibility of introducing a non-linear component into the model may help to solve this, for instance the second term in equation (1) representing the surface film formation.

$$R_c = \frac{\rho}{2a} + \frac{\rho_f d}{\pi a^2}$$

7.2 Data Driven Model Results

A prognostic model can be deemed to be acausal in its properties as its outputs and internal states depend upon future input values. One can attempt to define a model if operational conditions and load profiles are known, although as seen from the physics based model it is not always easy. To handle these uncertainties, Chapter 6 outlined a new type of data driven methodology that devises the model from the present and past operational and load profiles in order to infer a prediction of future events.

The results for the sliding window RLS derived model from Chapter 6 are shown below. The graphs feature the ARMA model upon the data, both in terms of the actual fit to the data and also the Mean Squared Error (MSE). This ability to know how accurate the model fit is to the present data alludes to a new 'on-line' prognostic metric. The MSE gives an indication of how good the prediction will be, and this can be seen in the results. Again, as before, the Prognostic Horizon (PH), alpha-lambda (α-λ) and Relative Accuracy (RA) metrics again are used to bench-mark the algorithm.

The difference in the window size to the overall prediction will play a part in the models accuracy, in the results, two window sizes are considered, 100 and 200 samples, respectively. It should be also noted that in the PH and α-λ plots there is an absence of data for the first 100 or 200 cycles (depending on the window), this is not a mistake,

the first prognosis may not be made until the window has made the first estimate based upon the window length.

The process of the estimation is illustrated in Figures 7.2 and 7.3, for measurements at 43000 and 83000 cycles respectively. Also, the estimated feature vector with error bounds that is extrapolated via the Kalman-filter from the ARMA model derived from the last L measurements within the window, until the threshold of failure is shown. It can be seen that the estimated feature vector has predicted the path of the measurements with a high degree of accuracy compared to the Particle filter, despite the variation of the data. Again the RUL is denoted as the distance between the current measurement and the failure threshold, a confidence boundary is present, although difficult to see due to the accuracy of the model.

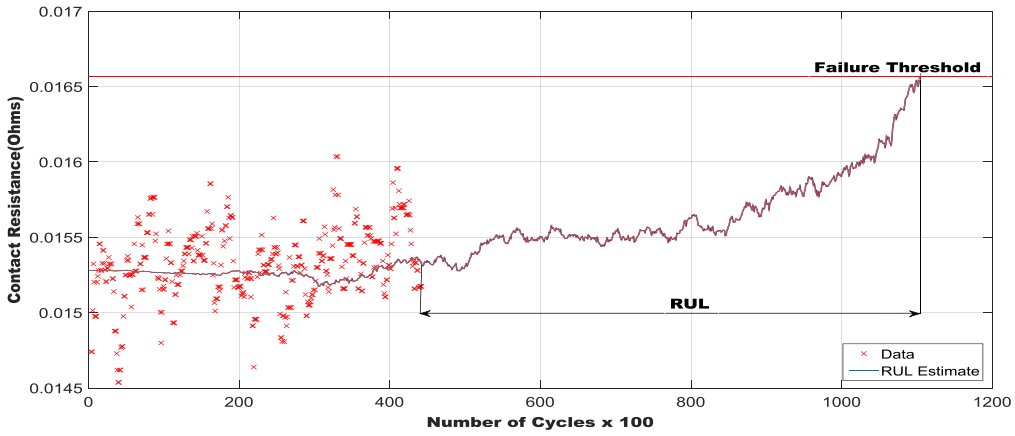


Figure 7.2. Remaining useful life after 40000 measurements

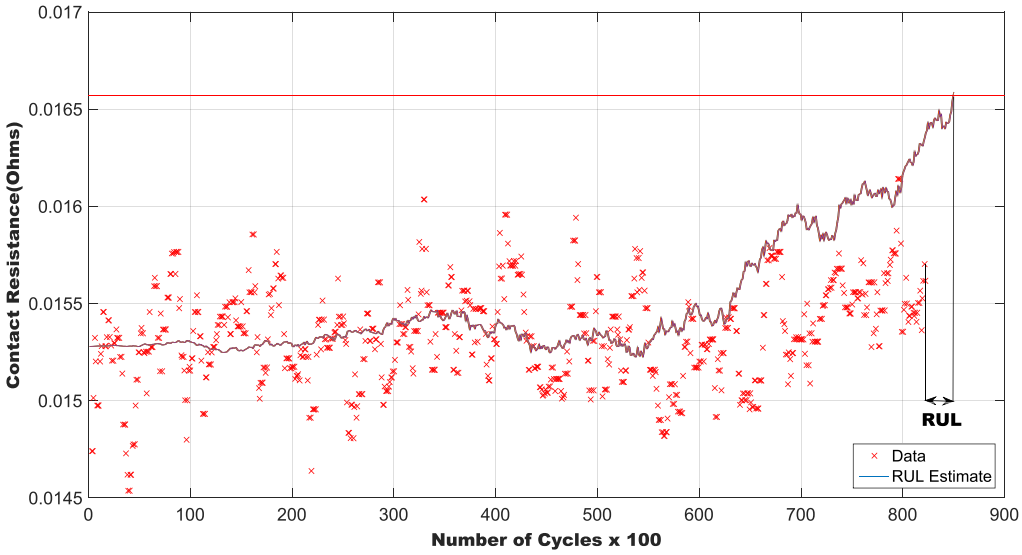


Figure 7.3. Remaining useful life after 80000 measurements

The RLS ARMA model for the six samples gives a promising result and are discussed briefly (the plots may be found in Chapter 6). Sample 1 is a complex data set, with large variations within the data and little upwards variation to the point of failure. This can be seen from the data and subsequent errors present in fitting the model to the data, errors in the order of 0.5 - $1x10^{-9}$, are present throughout. Around 700,000 cycles these errors manifest themselves in the RUL and α-λ plots as well as the RA prediction after 0.8λ. The RUL α-λ predictions are within 20% with a majority of the prediction being within 10%, with the RA being high initially, but falling off at 0.8λ. The window size being reduced by 50%, appears to have a small effect on the model, there is a noticeable reduction in the error plot which moves to 0.2 - 0.6 $x10^{-9}$ in magnitude. However, the overall predictions are unaltered, a still remain within 10% for the majority of the plots.

Sample 2 does not contain so much variation in the data and there is a clear increment to the threshold. The prediction as may be expected, is better. The overall error settles around $1x10^{-8}$, and there are less sharp spikes present, which is a reflection of the data. This can be seen in the metrics as well, the RUL and α-λ being within 10% and the RA is good until 0.8λ. Again the 100 sample window produces similar results, with little difference.

Sample 3 is again well defined and increases in a upward pattern towards the threshold, however, this time there is quite a lot of variation in the data, resulting in spikes of error around 1 - 4 $x10^{-9}$. Again the RUL and α-λ accuracy is quite good, there is significant amount of the prediction within 10% and the RA is good up to 0.9 λ.

Sample 4 again is made up of clearly defined, upwards trending data. One noticeable abnormality is the presence of a large spike at 79,000 cycles in the data which also clearly manifests itself in the error plot, however this does not impact on the prediction noticeably. The plots of RUL, α-λ are good for both 20% and 10% boundaries and RA is excellent until 0.9λ.

Sample 5 is of the nature of being a complex data set, considerable variation and only a slight upwards trending. Again, the error tended to lie around $0.5x10^{-8}$ in magnitude, with a degree of high peaks, which led to good plot results overall for the RUL and α-λ, with the 20% and 10% boundary achieved mostly throughout, however, the last 30,000 cycles were particularly full of peaks and trough which affected the accuracy.

Lastly, sample 6 was a clearly defined upward trending data set, with large swings in the data, this resulted in a model error of 0.2-0.8x10^{-8}, with sharp error peaks at around 7,000 and 80,000 cycles. The overall RUL and α-λ though was good, with the 20% boundary satisfied, and the majority of the prediction lying within the 10% boundary. The RA accuracy was good, being around 90% up till 0.9λ when the accuracy fell off sharply. Again the small window size, did not appear to effect the result greatly, with only a small gain in model error.

7.3 Discussion

The final section of this chapter summarises the prognostic results obtained from the two methodologies outlined in Chapters 5 and 6. The results from the physics of failure model are compared against the data driven solution in terms of their prognostic metrics. In the data driven model this is the RUL, alpha-lambda, RA and Model Error. For the physic model the RUL, alpha-lambda and RA are used respectively. Performance evaluation metric results are presented in the previous sections. In this section, the metric results are examined more thoroughly.

Table 7.1 below organises the prognostic results from each sample, in each case the average has been used to make comparison. Starting with the data driven approach, the sliding window RLS algorithm shows a good degree of accuracy with a window length of 100 samples achieving results in excess of 80% for most samples, samples 2 and 6 where the trending and variation in the data resulted in a poorer performance.

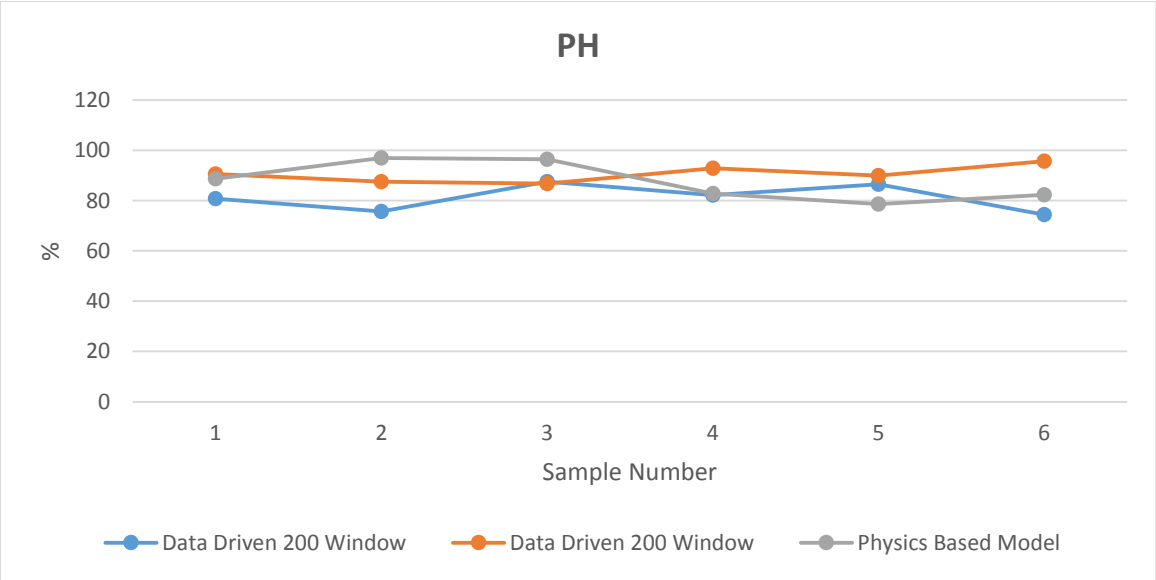
As the window is increased to 200 samples, the RUL results improve slightly as may be expected, moving towards 90% accuracy. This is also reflected in how the RA, α-λ (notes less is better in this metric) and RA improves and the noticeable drop in error.

In the physics based model, again the metrics are good, comparable in a majority of cases with the data driven model, which is surprising given the simplicity of the model approximation.

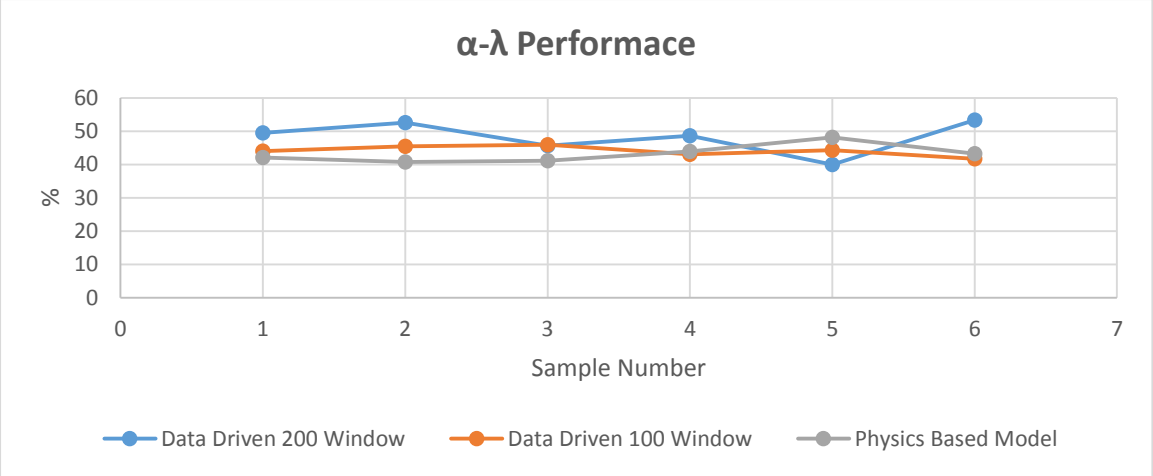
The graphs below illustrate the data in terms of the three main bench marks plotted together.

Data Driven 100 Samples Window	Sample1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
RUL(%)	80.76	75.63	87.53	82.13	86.56	74.43
20% α-λ (%)	49.5	52.61	45.64	48.63	40	53.36
10% α-λ (%)	24.73	26.3	22.82	24.29	23	26.68
RA(%)	83.95	93.68	91.07	95.44	94.74	92.6
Error(%)	1.87E-10	3.81E-10	5.17E-10	1.13E-09	8.69E-10	4.24E-10
Data Driven 200 Samples Window	**Sample1**	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**
RUL(%)	90.58	87.55	86.7	92.83	89.89	95.69
20% α-λ (%)	44	45.45	45.93	43.07	44.3	41.76
10% α-λ (%)	21.98	22.72	22.96	21.52	22.13	20.88
RA(%)	92.94	95.01	96.17	96.75	97.4	90.48
Error(%)	6.04E-10	1.02E-09	1.02E-09	1.89E-09	1.92E-09	8.59E-10
Physics Based Model	**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**
RUL(%)	88.64	97	96.43	82.76	78.64	82.36
20% α-λ (%)	42.14	40.81	41.16	43.91	48.22	43.25
RA(%)	94.46	99.7	97.43	90.31	82.11	89.9

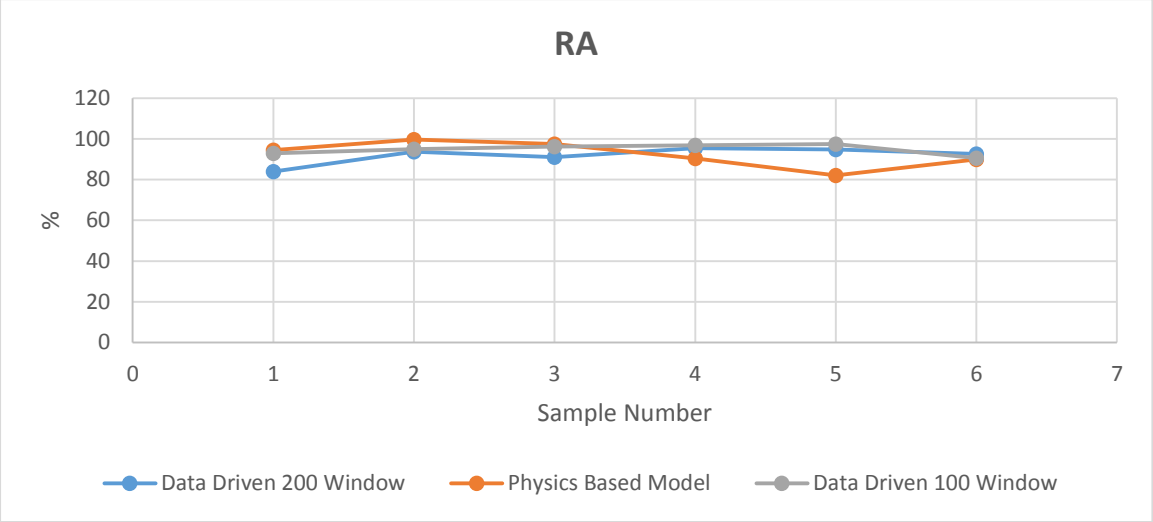
Table 7.1 Comparison of the two methodologies namely Physics based versus the Data Driven approach.



7.4 Comparison of the Prognostic Horizon results



7.5 Comparison of the α-λ results.



7.6 Comparison of the RA results.

7.3.1 Computational Execution Time

One important consideration of the algorithms developed is the execution time. This was timed for a number of iterations on both the sliding window RLS algorithm utilising the Kalman filter (200 sample window length) and the PoF Model with the Particle filtering (3000 particles) algorithms respectively, within the Matlab™ platform.

Algorithm	Average Execution Time (s)
Sliding window RLS with Kalman filter	1.1
PoF based with Particle filter	29.9

Table 7.2 Comparison of algorithm execution times

7.3.2 Conclusion

In conclusion, the main goal of this research was to look at the development of a computationally efficient prognostic methodology that allowed the estimation of relay life in real time. From the work carried out, two solutions have been proposed, both performing the task set out to a high degree of accuracy. However, both methods have associated with them pros and cons. In the PbM, the model first needs to be set up, which involves investigating the physical parameters such as mass loss from the arcing duration and this is set within the model and hence the model is not adaptable if say the loading changed. Another downside to the PbM is the computational time, however due to the model being linear, this could possibly be shortened by the use of a Kalman Filter for the prediction.

On the other hand, the data driven solution is adaptable, as it uses past data within a window for the future life prediction and is therefore self-training, hence it is able to adapt to parameter changes. As well adaptation, it is also computationally efficient and can easily be operated in real time within a DSP platform for example. Another advantage is the ability to respond to servicing, if for example a component is changed, within the window size of samples, a new model of the degradation may be built. Future work also needs to be carried out in order to enable the optimisation of the window size.

Chapter 8 - Conclusions and Future work

IVHM is an enabling technology that is starting to gain ground amongst operators of critical assets such as aerospace equipment, where maintenance is a large expenditure for both the operator and the manufacturer. The use of prognostics fits into the IVHM framework of system monitoring, fault detection and diagnosis, failure prognosis and operations management. Prognostics is the newcomer to this framework, with fault detection and diagnosis being a mature technology relatively speaking. However, an intelligent prognostics capability implemented such that the maintainers are able to predict failures and the remaining service life of these systems and associated components is a very attractive technology. The long term aim to build prognostics capability into equipment such that cost-effective, safe, reliable and robust systems can be built with extended service life has been comprehended both by industry and the research communities.

The overall aim of prognostics is to provide an accurate estimation of the remaining useful life of an asset. The prognosis also needs to take into account the operating conditions and how the system is to be used in the future in order to give a RUL prediction and the user beneficial information that allows decisions to be made based on safety and cost, depending on the priority of the operational use.

Prognostics may fall into three main categories, namely experience based, Data Driven and Physics of Failure. Cost and accuracy drive the type of model selection. Experienced based prognostics is the cheapest option and its range of system applicability is diverse, however in terms of accuracy, it falls short of physics of failure based models which are the most accurate but difficult and costly to develop. Data Driven based methodologies form the middle ground, however model-based prognostics, are expected to take prognostics into the future.

In this thesis, the development of both a physics based model and data driven solution are explored and both methodologies are explored in terms of their suitability for the assessment of the intended asset, the relay outputs on the Full Authority Digital Controller (FADEC) for a range of aircraft currently being developed.

Electrical wear of the contacts was identified as the primary mode of failure in the relays, due to Joule heating and erosion of the surface to arcing, contamination and corrosion. The constriction (contact) resistance was deemed to be an important metric

in determining how the condition of the contact may be determined, as well as its importance in the development of failure. The complex nature of the formation of an arc between the contacts was examined and how this erodes the surface of the contact due to material loss and mass transfer. The process was determined as being a complex multi-faceted process and influences the prediction of how long the relay contacts will last.

The contribution of this research was to examine the feasibility of condition monitoring and prognostic implementation for the relays. Experimental testing to enable real world data to be collected and thus replicate the failure mode was carried out. A suitable feature vector in the form of the contact resistance was determined from this experimental work to enable the RUL to be predicted. From the experimental work the development and evaluation of suitable prognostic algorithms to handle the complex degradation process from the data collected was carried out. Finally, the development of tailored, real time prognostic algorithms that are implementable at system level was trialled and benchmarked to assess their performance.

Two methods were developed based upon a physic based solution. Firstly, a simple physics model based upon arc erosion of the contact and its effect on the contact resistance due to the reduction of contact surface area was tried. The physics based model of the arcing and the erosion process was developed to enable mass loss to be calculated, the results were used in a state space model. This model was then used with Particle filtering to enable a prediction of the RUL to be estimated. Secondly, a computationally efficient, technique based upon system identification using Recursive Least Squares algorithm via a sliding window was developed to produce the degradation model which was used in conjunction with a Bayesian framework in the form of a Kalman-filter for state variable evolution. Remaining useful life was calculated based on the evolution of the state space feature vector.

Both the models were evaluated against proven prognostic metrics from literature used to bench mark the effectiveness of algorithms. The results from both models were consistently within 20% accuracy and most of the time in excess of this, leading to the conclusion that the prognostic performance results show that the both physics-based and data-driven models predict the system behaviour accurately which led to successful RUL predictions.

The major downfall with the physics model was that any alterations to the model material loss parameter needed to be gained from the arc model and heat equation, calculated separately, hence the model is not a true physics based model in the sense that it may adapt to changes in conditions such as loading, environmental or material deviations. Other flaws included the considerable computation power needed when using the Particle filter, to get a reasonable estimate, this evoked a great deal of processing time even on a Core™ i5 machine.

On the other hand, the data driven solution is adaptable, as it uses past data within a window for the future life prediction and is therefore self-training, hence it is able to adapt to parameter changes. As well as adaptation, it is also computationally efficient and can easily be operated in real time within a DSP platform for example. Another advantage is the ability to respond to servicing, if for example, a component is changed within the window size of samples, a new model of the degradation may be built.

The key areas for future work will be to investigate further the development of these models and how factoring additional failure parameter may increase the accuracy. This will involve the extraction of a secondary feature vector from the test rig, such as contact closure/release time, to see how this may influence the data driven model. As well as this, optimisation of the window size needs to be carried out. One interesting aside that also arises due to this model was the ability to have an idea of what the model error was before the prediction, which leads to the feasibility of devising an on-line metric for RUL accuracy.

Whereas for the physics based model, bringing additional parameters into the model such as material transfer or the effects of film formation, either as empiric factors or modelling would be of interest.

References

Ak, R., Li, Y. F., Vitelli, V. and Zio, E. (2013), "A Genetic Algorithm and Neural Network Technique for Predicting Wind Power under Uncertainty ", *Chemical Engineering Transactions*, vol. 33, pp. 925-930.

Alcock, Charles B., "*Thermochemical processes: principles and models*", Butterworth-Heinemann, 2000.

Alghassi, Alireza, et al. "Computationally efficient, real-time, and embeddable prognostic techniques for power electronics." *Power Electronics, IEEE Transactions on* 30.5 (2015): 2623-2634.

Amin, Sanket, Carl Byington, and Matthew Watson. "Fuzzy inference and fusion for health state diagnosis of hydraulic pumps and motors." *Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American*. IEEE, 2005.

Ammerman, Ravel F., et al. "DC-arc models and incident-energy calculations." *IEEE Transactions on Industry Applications* 46.5 (2010): 1810-1819.

An, Dawn, Joo-Ho Choi, and Nam Ho Kim. "Prognostics 101: A tutorial for particle filter-based prognostics algorithm using matlab." *Reliability Engineering & System Safety* 115 (2013): 161-169.

An, Dawn, Joo-Ho Choi, and Nam Ho Kim. "*Prognostics 101: A tutorial for particle filter-based prognostics algorithm using matlab*." Reliability Engineering & System Safety 115 (2013): 161-169.

Antler, M. and Sproles, E. S., IEEE Trans. CHMT, 5(1), 58–166, 1982.

Artes, M., Del Castillo, L., Perez, J., "Failure prevention and diagnosis in machine elements using cluster", *Proceedings of the Tenth International Congress on Sound and Vibration*, Stockholm, Sweden, 2003, pp. 1197–1203.

Arulampalam, M. S, Maskell, S., Gordon, N., and Clapp, T. "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking", *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp.174–188, 2002.

Astrom, K. J. and Wittenmark, B., "*Adaptive Control*," Addison-Wesley Publishing, 1995.

Ayrton, H., "The Electric Arc", London, U.K.: Electrician, 1902.

Balme, W., P. Braumann, and K. H. Schröder. "Welding tendency, contact resistance and erosion of contact materials in motor vehicle relays." *Electrical Contacts, 1990. Proceedings of the Thirty-Sixth IEEE Holm Conference on... and the Fifteenth International Conference on Electrical Contacts*. IEEE, 1990.

Baraldi, P., Mangili, F. and Zio, E. (2015), "A prognostics approach to nuclear component degradation modeling based on Gaussian Process Regression", *Progress in Nuclear Energy*, vol. 78, no. 0, pp. 141-154.
Baraldi, Piero, Francesca Mangili, and Enrico Zio. "A kalman filter-based ensemble approach with application to turbine creep prognostics." *Reliability, IEEE Transactions on* 61.4 (2012): 966-977.
Belge M and Miller E L. "A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise." IEEE Signal Processing Letters, 7(4): 86 -89. 2000.

Biagetti, T., Sciubba, E., "Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems", *Energy 29* (12–15) (2004) 2553–2572.
Box, G.E.P., Jenkins, G.M., "*Time Series Analysis: Forecasting and Control*", Holden-Day, San Francisco, 1976.
Braunovic´, M., IEEE Trans. CHMT, 15, 204–214, 1992.

Braunovic, Myshkin, and Konchits, "*Electrical contacts: fundamentals, applications and technology"*, CRC press, 2006.
Brotherton, T., Jahns, G., Jacobs, J. and Wroblewski, D. (2000), "Prognosis of faults in gas turbine engines", *IEEE Aerospace Conference Proceedings*, Vol. 6, pp. 163.

Brown, P. and Sondalini, M., "The Evolution of Maintenance Practices*"*, available at: http://www.lifetime-reliability.com/free-articles/maintenancemanagement/Evolution_of_Maintenance_Practices.pdf (accessed 05/3/2016).,2014.

Bunks, Carey, Dan McCarthy, and Tarik Al-Ani. "Condition-based maintenance of machines using hidden Markov models." *Mechanical Systems and Signal Processing* 14.4 (2000): 597-612.

Byer, B., Hess, A., and Fila, L. Writing a convincing cost benefit analysis to substantiate autonomic logistics. In Proceedings of the 2001 IEEE Aerospace Conference, Big Sky, Montana, USA, 10–17 March 2001, vol. 6, pp. 3095– 3103.

Byington, Carl S., Matthew Watson, and Doug Edwards. "Data-driven neural network methodology to remaining life predictions for aircraft actuator components." *Aerospace Conference, 2004. Proceedings. 2004 IEEE*. Vol. 6. IEEE, 2004.

Carden, E. Peter, and Paul Fanning. "Vibration based condition monitoring: a review." *Structural health monitoring* 3.4 (2004): 355-377.

Chen, C., Vachtsevanos, G. and Orchard, M., "Machine remaining useful life prediction: An integrated adaptive neuro-fuzzy and high-order particle filtering approach," *Mechanical Syst. and Signal Process.*, vol. 28, pp. 597–607, 2011.

Chen, C., Zhang, B., Vachtsevanos, G. and Orchard, M., “Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering,” *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4353–4364, 2011.

Chen, Peng, et al. "Research on dynamic contact resistance measurement and failure prediction of relay contacts." *Advanced Technology of Electrical Engineering and Energy* 24.3 (2005): 27.

Chen, Z. K., and G. J. Witter. "A Study of Dynamic Welding of Electrical Contacts with Emphasis on the Effects of Oxide Content for Silver Tin Indium Oxide Contacts." *2010 Proceedings of the 56th IEEE Holm Conference on Electrical Contacts*. IEEE, 2010.

Chen, Zhuan-Ke, and Koichiro Sawa. "The investigation of surface characteristics and contact resistance of DC relay contacts." *Components, Hybrids, and Manufacturing Technology, IEEE Transactions on* 16.2 (1993): 211-219.

Cheng, Bing, and D. Michael Titterington. "Neural networks: A review from a statistical perspective." *Statistical science* (1994): 2-30.

Chinnam, Ratna Babu, and Pundarikaksha Baruah. "A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems." *International Journal of Materials and Product Technology* 20.1-3 (2004): 166-179.

Choi, Seong Soo, and Soon Heung Chang. "Development of an on-line fuzzy expert system for integrated alarm processing in nuclear power plants." *Nuclear Science, IEEE Transactions on* 42.4 (1995): 1406-1418.

Coppe, Alexandra, et al. "Reducing uncertainty in damage growth properties by structural health monitoring." *Annual Conference of the Prognostics and Health Management Society*. Vol. 27. 2009.

Daigle, M., and Goebel, K., “Improving Computational Efficiency of Prediction in Model-Based Prognostics using the Unscented Transform,” Tech. Rep. DTIC Document, 2010.

Daigle, M. and Goebel, K., “Model-based prognostics under limited sensing,” in *Proc. IEEE Aerospace Conf.*, 2010, pp. 1–12.

Davies, C., Greenough R.M., “The use of information systems in fault diagnosis”, Proceedings of the 16th National Conference on Manufacturing Research, University of East London, UK, 2000.

Deng, Julong, "Introduction to grey system theory." *The Journal of grey system* 1.1 (1989): 1-24.

Deng, Ju-Long, "Control problems of grey systems." *Systems & Control Letters* 1.5 (1982): 288-294.

Diesselhorst, H., “Uber das Problem eines elektrisch erwarmten Leiters”, *Ann Phys* Leipzig 1:312, 1900.

Dong, Y., Gu, Y., Yang, K. and Zhang, W., "A combining condition prediction model and its application in power plant", Proceedings of 2004 International Conference on Machine Learning and Cybernetics, Vol. 6, pp. 3474. (2004)

Dron, J.P., Rasolofondraibe, L., Couet, C., Pavan, A., “Fault detection and monitoring of a ball bearing benchtest and a production machine via autoregressive spectrum analysis”, *Journal of Sound and Vibration* 218 (1998) 501–525.

Dutton, Ken, Steve Thompson, and Bill Barraclough, “*The art of control engineering*”, Addison-Wesley Longman Publishing Co., Inc., 1997.

Ebling, W., Foster, A., Radtke, R., “Physics of Non-Idealized Plasmas*”, Stuttgart: Teubner Verlagsgesellschaft*, 1991.

Elliott, R.J., Aggoun, L., Moore, J.B., “*Hidden Markov Models: Estimation and Control*”, Springer, New York, 1995.

Fan, Jiajie, K. C. Yung, and Michael Pecht. "Physics-of-failure-based prognostics and health management for high-power white light-emitting diode lighting." *Device and Materials Reliability, IEEE Transactions on* 11.3 (2011): 407-416.

Fang Y., Zhigang, L.,Yachao, W.,Cuiping, Y., “Degradation Failure Model of Electromagnetic Relay”, *26th International Conference on Electrical Contacts*, pages 116 –123. (2012).

Feng, Y., Lul, J., Zheng, J., Huang,Z., “Research on the Failure Diagnostics Parameters and the Reliability Prediction Model of the Electrical Contacts”, Electrical contacts - 2006, proceedings of the fifty-second IEEE Holm Conference on Electrical Contacts, pages 69-72., (2006).

Fujitsu Components Engineering Reference: Relays 2009, Fujitsu Components America.

Gammon, T. and Matthews, J., “Conventional and recommended arc power and energy calculations and arc damage assessment,” *IEEE Trans. Ind. Appl.* , vol. 39, no. 3, pp. 594–599, May/Jun. 2003.

Garga, A., McClintic, K., Campbell, R., Yang, C. Lebold, M. Hay, T. and Byington, C., “Hybrid reasoning for prognostic learning in cbm systems,” in *Proc. IEE Aerospace Conf.*, 2001, vol. 6, pp. 2957–2969.

Goebel, K. Eklund, N. and Bonanni, P., "Prognostic Fusion for Uncertainty Reduction", *Wright-Patterson AFB*, OH, USA: Defense Technical Information Center, 2007.

Goebel, K., Eklund, N., and Bonanni, P., "Fusing competing prediction algorithms for prognostics," in *Proc. IEEE Aerospace Conf.*, 2006, p. 10.

Gola, G. and Nystad, B., "From measurement collection to remaining useful life estimation: Defining a diagnostic-prognostic frame for optimal maintenance scheduling of choke valves undergoing erosion," in *Proc. Annu. Conf. Prognost. Health Management Soc.*, 2011, pp.26–29.

Goode, K., Moore, J. and Roy lance, B., "Plant machinery working life prediction method utilizing reliability and condition-monitoring data," in *Proc. Institute. Mechanical. Engineers, Part E: J. Process Mechanical Eng.*, 2000, vol. 214, pp. 109–122, 2.

Greenwood, J. A., and J. B. P. Williamson. "Electrical conduction in solids. II. Theory of temperature-dependent conductors." *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 246. No. 1244. The Royal Society, 1958.

Greenwood, J. A., "Constriction resistance and real area of contact," Brit. J. Appl. Phys., Vol.17, pp.162-1632, 1966.

Grewal, Mohinder S. "*Kalman filtering*", Springer Berlin Heidelberg, 2011.

Groer, P. G., ''Analysis of Time-to-Failure with a Weibull Model,'' *Proceedings of the Maintenance and Reliability Conference*, MARCON 2000, Knoxville, TN.

Gu, Jie, et al. "Application of grey prediction model for failure prognostics of electronics." *International Journal of Performability Engineering* 6.5 (2010): 435-442.

Guclu, Adem, et al. "Prognostics with autoregressive moving average for railway turnouts." *Annual Conference of Prognostics*. 2010.

Guo, H., Watson, S., Tavner, P. and Xiang, J., "Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation," *Rel. Eng. Syst. Safety*, vol. 94, no. 6, pp.1057–1063, 2009.

Hall, P. M., Myers, K. and Vilcheck, S. W., "Arcing faults on direct current trolley systems", *Proc. 50th WVU Conf. Coal Mine Electrotechnol.,* Morgantown, WV, 1978, pp. 1–19.

Hammerschmidt, Martin, Alexander R. Neuhaus, and Werner F. Rieder. "The effects of material transfer in relays diagnosed by force and/or voltage measurement." *Components and Packaging Technologies, IEEE Transactions on* 27.1 (2004): 12-18.

Haug, R., Kouakou, T., Doremieux, J., “Phenomena proceeding arc ignition between opening contacts: experimental study and theoretical approach”, *Proceedings of the 36th IEEE Holm Conference on Electrical Contacts:* 543–549, 1990.

Haykin, “*Adaptive Filter Theory*,” 5th edition, Prentice Hall. 2013.

Hecht, Herbert. "Why prognostics for avionics?" *Aerospace Conference, 2006 IEEE*. IEEE, 2006.

Hess, A., Calvello, G., and Dabney, T. PHM a key enabler for the JSF autonomic logistics support concept. In Proceedings of the 2004 IEEE Aerospace Conference, Big Sky, Montana, USA, 6–13 March 2004, vol. 6, pp. 3543–3550.

Holm, R, “*Electric Contacts, Theory and Applications”*, Berlin: Springer–Verlag, 2000.

Huang, R. Xi, L. Li, X., Richard Liu, C., Qiu, H. and Lee, J., “Residual life predictions for ball bearings based on self-organizingmap and back propagation neural network methods,” *Mechan. Syst. Signal Process.*, vol. 21, no. 1, pp. 193–207, 2007.

Huang, Runqing, et al. "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods." *Mechanical Systems and Signal Processing* 21.1 (2007): 193-207.

Hwang, P. Y. C., Brown, R. G. Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions. *John Wiley Sons, Inc.*, fourth edition, 2012.

Isermann, Rolf. “*Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*”. Springer Science & Business Media, 2011.

Ishida, H., Watanabe, Y., Taniguchi, M., Inoue, H., Takago, T., “Observation of contact bridge phenomena at transient and steady state”, *Proceedings of the 50th IEEE Holm Conference on Electrical Contacts:* 519–522, 2004.

ISO 13381-1, Condition Monitoring and Diagnostics of Machines – Prognostics – Part 1: General Guidelines: International Standards Organization, 2004.

James, Glyn. "Advanced Modern Engineering Mathematics”, 2011.

Janasak, K. M. and Beshears, R. R. Diagnostics to prognostics – a product availability technology evolution. In Proceedings of the 2007 Reliability and Maintainability Symposium – RAMS'07, Orlando, Florida, USA, 22–25 January 2007, pp. 113–118.

Jardine, Andrew KS, Daming Lin, and Dragan Banjevic. "A review on machinery diagnostics and prognostics implementing condition-based maintenance." *Mechanical systems and signal processing* 20.7 (2006): 1483-1510.

Jemaa, N. Ben, L. Nedelec, and S. Benhenda. "Break arc duration and contact erosion in automotive application." *Components, Packaging, and Manufacturing Technology, Part A, IEEE Transactions on* 19.1 (1996): 82-86.

Jennions, I. K., "Integrated Vehicle Health Management: Perspectives on an Emerging Field", SAE International, 2011.

Jiang, J. and Zhang, Y., "A novel variable-length sliding window blockwise least-squares algorithm for on-line estimation of time-varying parameters", *Int. J. Adapt. Control Signal Processing*, 18: 505–521. doi: 10.1002/acs.811, 2004.

Khan, Faisal, et al. "Prognostics of crack propagation in structures using time delay neural network." *Prognostics and Health Management (PHM), 2015 IEEE Conference on*. IEEE, 2015.

Kim, Phil., "*Kalman filter for beginners: with MATLAB examples*.", Vol. 4. CreateSpace, 2011.

Kirianaki, N.V. Yurish, S.Y. Shpak, N.O. Deynega, V.P. "*Data Acquisition and Signal Processing for Smart Sensors*", Wiley, Chichester, West Sussex, England, 2002.

Kohlrausch, F., "Uber den Stationaren Temperaturzustand eines elektrisch Geheizten Leiters", *Ann Phys*, Leipzig 1: 132, 1900.

Koren, P.P., Nahemow, M.D., Slade, P.G., "The molten metal bridge state of opening electric contacts", *IEEE Trans. Parts Hyb. Pack.* PHP–11(1): 4–10, 1975.

Kothamasu, R., Huang, S. H. and Verduin, W. H., "System health monitoring and prognostics - A review of current paradigms and practices", International Journal of Advanced Manufacturing Technology, vol. 28, no. 9, pp. 1012-1024, 2006.

Kulkarni, C. S., Celaya, J. R., Goebel, K. and Biswas, G., "Bayesian framework approach for prognostic studies in electrolytic capacitor under thermal overstress conditions,*"Annual Conference of the Prognostics and Health Management Society*, 2012.

Kulkarni, Chetan S., et al. "Physics based Modeling and Prognostics of Electrolytic Capacitors." *Proc. AIAA* (2012): 1-14.

Kuniewslci, S.P., van der Weide, J.A.M., van Noortwijk, J.M., “Sampling inspection for the evaluation of time-dependent reliability of deteriorating systems under imperfect defect detection”, *Reliability Engineering and System Safety 94* (9), 1480–1490. 2009.

Lall, P., Lowe, R., Goebel, K., “Prognostics health management of electronic systems under mechanical shock and vibration using Kalman filter models and metrics”, *Industrial Electronics, IEEE Transactions on 59 (11*), 4301-4314, 2012.

Lall, Pradeep, et al. "Leading indicators for prognostic health management of electrical connectors subjected to random vibration." *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2012 13th IEEE Intersociety Conference on*. IEEE, 2012.

Lall, Pradeep, Ryan Lowe, and Kai Goebel. "Extended Kalman filter models and resistance spectroscopy for prognostication and health monitoring of leadfree electronics under vibration." *Reliability, IEEE Transactions on* 61.4 (2012): 858-871.

Lall, Pradeep, Ryan Lowe, and Kai Goebel. "Extended Kalman filter models and resistance spectroscopy for prognostication and health monitoring of leadfree electronics under vibration." *IEEE Transactions on Reliability* 61.4 (2012): 858-871.

Lall, Pradeep, Ryan Lowe, and Kai Goebel. *"*Extended Kalman Filter Models and Resistance Spectroscopy for Prognostication and Health Monitoring of Leadfree Electronics Under Vibration." *IEEE TRANSACTIONS ON RELIABILITY* 61.4, 2012.

Lall, Pradeep, Ryan Lowe, and Kai Goebel. "Particle filter models and phase sensitive detection for prognostication and health monitoring of lead free electronics under shock and vibration." *Electronic Components and Technology Conference (ECTC), 2011 IEEE 61st*. IEEE, 2011.

Lall, Pradeep, Ryan Lowe, and Kai Goebel. "Prognostics using Kalman-Filter models and metrics for risk assessment in BGAs under shock and vibration loads." *Electronic Components and Technology Conference (ECTC), 2010 Proceedings 60th*. IEEE, 2010.

Lawless, J., Crowder, M., “Covariates and random effects in a Gamma process model with application to degradation and failure”, *Lifetime Data Analysis 10*, 213–227, 2004.

Lee, J., Abujamra, R., Jardine, A.K.S. , Lin, D., Banjevic, D., “An integrated platform for diagnostics, prognostics and maintenance optimization”, *The IMS ’2004 International Conference on Advances in*

Maintenance and in Modeling, Simulation and Intelligent Monitoring of Degradations, Arles, France, 2004.

Lee, Jong Min, et al. "Diagnosis of mechanical fault signals using continuous hidden Markov model." *Journal of Sound and Vibration* 276.3 (2004): 1065-1080.

Lee, Jong-Min, ChangKyoo Yoo, and In-Beum Lee. "Statistical process monitoring with independent component analysis." *Journal of Process Control* 14.5 (2004): 467-485.

Lee, A., Mao, A., and Mamrick, M. S., *Proceedings of 34th IEEE Holm Conference on Electrical Contacts*, San Francisco, 87–91, 1988.

Leung, C., Streicher, E., Fitzgerald, D., Cook, J., "High current erosion of Ag-SnO2 contacts and the evaluation of indium effects in oxide properties", *Proc 52nd IEEE Holm Conference Electrical Contacts* 143–150, 2006.

Leung, Chi H., and Anthony Lee. "Contact erosion in automotive DC relays." *Components, Hybrids, and Manufacturing Technology, IEEE Transactions on* 14.1 (1991): 101-108.

Li, Kui, et al. "Test and analysis of reliability for electromagnetic relay." *Electrical Contacts, 2000. Proceedings of the Forty-Sixth IEEE Holm Conference on*. IEEE, 2000.

Liao, L. and Kottig, F. "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction", *IEEE Trans. Rel*., vol.63, no.1, pp.191–207, 2014.

Liao, Linxia, and Felix Kottig. "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction." *Reliability, IEEE Transactions on* 63.1 (2014): 191-207.

Liu, J., Wang, W., Ma, F., Yang, Y. and Yang, C., "A data-model-fusion prognostic framework for dynamic system state forecasting," *Engin. Appl. Artif. Intell.*, vol. 25, pp. 814–823, 2012.

Liu, Z. Li, Q. and Mu, C., "A hybrid LSSVR-HMM based prognostics approach," in *Proc. 4th Int. Conf. Intell. Human-Machine Syst. Cybernet.*, 2012, vol. 2, pp. 275–278.

Ljung, L. "*System Identification: Theory for the User, 2nd ed.",* New Jersey: Prentice-Hall, 1999.

Lopez, Leoncio D., and Michael G. Pecht. "Maxima-SPRT methodology for health monitoring of contact resistance in IC sockets." *Prognostics and Health Management, 2008. PHM 2008. International Conference on*. IEEE, 2008.

M.S. Nixon, A.S. Aguado, "*Feature Extraction and Image Processing*", Newnes, Oxford, 2002.

Mahamad, A. K., Saon, S. and Hiyama, T. (2010), "Predicting remaining useful life of rotating machinery based artificial neural network", *Computers & Mathematics with Applications*, vol. 60, no. 4, pp. 1078-1087.

Malhi, Arnaz, Ruqiang Yan, and Robert X. Gao., "Prognosis of defect propagation based on recurrent neural networks." *Instrumentation and Measurement, IEEE Transactions on* 60.3 (2011): 703-711.

Marafioti G. Bitmead R. R. *and* Hovd M.*, "*Persistently exciting model predictive control*," Int. J. Adapt. Control Signal Process*., 28*, pages* 536–552. *2014*.

McBride, J.W., Jiang, L., Chianrabutra, C., "Fine transfer in electrical switching contacts using gold coated carbon nano-tubes", *Proceedings of the 26th International Conference on Electrical Contact Phenomena:* 353–358, 2012.

McBride, John W. "An experimental investigation of contact bounce in medium duty contacts." *IEEE transactions on components, hybrids, and manufacturing technology* 14.2 (1991): 319-326.

McBride, John W. "Electrical contact bounce in medium-duty contacts." *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* 12.1 (1989): 82-90.

McBride, John W. "The volumetric erosion of electrical contacts." *IEEE Transactions on Components and Packaging Technologies* 23.2 (2000): 211-221.

McManus, Hugh, and Daniel Hastings, "A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems", *INCOSE International Symposium*. Vol. 15. No. 1. 2005.

Miller, D. B. and Hildenbrand, J. L., "DC arc model including circuit constraints," *IEEE Trans. Power App. Syst.,* vol. PAS-92, no. 6, pp. 1926–1934, Nov. 1973.

Miyajima, K., Nitta,S., Mutoh, A., "A proposal on contact surface model of electro-magnetic relays—based on the change of showering arc wave formations with the number of contact operations", *IEICE Trans. Electron*. E81-C(3): 399–407, 1998.

Morin, Laurent, et al. "Contacts materials performances under break arc in automotive applications." *Components and Packaging Technologies, IEEE Transactions on* 23.2 (2000): 367-375.

Morin, Laurent, N. B. Jeman, and Didier Jeannot. "Make arc erosion and welding in the automotive area." *Components and Packaging Technologies, IEEE Transactions on* 23.2 (2000): 240-246.

Mulgrew, Bernard, and Colin F. Cowan. "*Adaptive filters and equalisers."* Vol. 56. Springer Science & Business Media, 2012.

Nandi, Subhasis, Hamid A. Toliyat, and Xiaodong Li. "Condition monitoring and fault diagnosis of electrical motors-a review." *Energy Conversion, IEEE Transactions on* 20.4 (2005): 719-729.

NASA, "https://www.nasa.gov/directorates/heo/scan/engineering/technology/txt_accordion1.html", *2015*.

Nelson, Wayne. "Accelerated life testing-step-stress models and data analyses." *Reliability, IEEE Transactions on* 29.2 (1980): 103-108.

Nottingham, W. B., "A new equation for the static characteristic of the normal electric arc", *Trans. Amer. Inst. Elect. Eng.*, vol. 42, p. 302, 1923.

Nottingham, W. B., "Normal arc characteristic curves: Dependence on absolute temperature of anode*,"* *Phys. Rev.*, vol. 28, no. 4, pp. 764–768, Oct. 1926.

Omron Components, Relay User's Guide,1990. https://www.components.omron.com/components/web/pdflib.nsf/0/3B7C5B7BB9EA88838525720100 7DD66D/$file/M20Z54E11.pdf

Ompusunggu, Agusmian, Jean-Michel Papy, and Steve Vandenplas. "Kalman filtering based prognostics for automatic transmission clutches ", *Mechatronics, IEEE/ASME Transactions on* (Volume:PP, Issue: 99), 2015

Orchard ME, Vachtsevanos GJ. "A particle-filtering approach for on-line fault diagnosis and failure prognosis", *Transactions of the Institute of Measurement and Control*, 31(3–4):221–46, (2009)

Orchard, M. E. and Vachtsevanos, G.), "A particle filtering approach for on-line failure prognosis in a planetary carrier plate", International Journal of Fuzzy Logic and Intelligent Systems, vol. 7, no. 4, pp. 221-227, (2007)

Orchard, Marcos E., and George J. Vachtsevanos. "A particle-filtering approach for on-line fault diagnosis and failure prognosis." *Transactions of the Institute of Measurement and Control* (2009).

Orchard, Marcos, Biqing Wu, and George Vachtsevanos. "A particle filtering framework for failure prognosis." *World tribology congress III*. American Society of Mechanical Engineers, 2005.

Orchard, Marcos, et al. "Advances in uncertainty representation and management for particle filtering applied to prognostics." *Prognostics and health management, 2008. phm 2008. International conference on*. IEEE, 2008.

Orsagh, R., Sheldon, J. and Klenke, C. "Prognostics/diagnostics for gas turbine engine bearings," in *Proc. IEEE Aerospace Conf.*, 2003, pp. 1165–1173.

Ostendorf, Frank, Thomas Wielsch, and Michael Reiniger. "Reliability assessment and field failure predictions-a prognostic model for separable electrical contacts." *ICEC 2014; The 27th International Conference on Electrical Contacts; Proceedings of*. VDE, 2014.

Peel, Leto. "Data driven prognostics using a Kalman filter ensemble of neural network models." *Prognostics and Health Management, 2008. PHM 2008. International Conference on*. IEEE, 2008.

Peng, T., He, J., Liu, Y., Saxena, A., Celaya, J. and Goebel, K., Integrated fatigue damage diagnosis and prognosis under uncertainties, *National Aeronautics And Space Administration Moffett Field CA AMES Research Center*, (2012).

Peng, Z. K., and F. L. Chu. "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography." *Mechanical systems and signal processing* 18.2 (2004): 199-221.

Pham, Hong Thom, and Bo-Suk Yang. "Estimation and forecasting of machine health condition using ARMA/GARCH model*." Mechanical Systems and Signal Processing 24.2 (2010): 546-558*.

Proakis, John G. "Digital signal processing: principles, algorithms, and application-3/E.", Prentice Hall, (1996).

Puchkarev, V., Bochkarev, M., "High current density spotless vacuum arc as a glow discharge", *IEEE Trans. Plasma Sci.* 25(4): 593–597, 1997.

Qiong, Y.,Xuerong, Y,.Guofu, Z., "Reliability Assessment for Electromagnetic Relay Based on Time Parameters Degradation", 11th International Conference on Electronic Packaging Technology & High Density Packaging (ICEPT-HDP), pages 1269 -1272.,(2010).

Rabiner, L.R., "Tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE 77* (1989) 257–286.

Rafiee, J., et al. "Intelligent condition monitoring of a gearbox using artificial neural network." *Mechanical systems and signal processing* 21.4 (2007): 1746-1754.

Rieder, W., Weichsler, V.,"Make erosion mechanism of AgCdO and AgSnO2 contacts", *IEEE Trans Components, Hybrids Manuf Technol* 15 (3): 332–338, 1992.

Rosunally, Y. Z., Stoyanov, S., Bailey, C., Mason, P., Campbell, S. and Monger, G. (2009), "Prognostics framework for remaining life rediction of cutty sark iron structures", *Annual Conference of the Prognostics and Health Management Society 2009*, 2009, San Diego, CA, USA, PHM Society, San Diego, CA, USA.

Saha, B., Goebel, K., Poll, S. and Christophersen, J. (2007), "An integrated approach to battery health monitoring using bayesian regression and state estimation", *Autotestcon*, 2007 IEEE, pp. 646.

Saha, Bhaskar, and Kai Goebel. "*Model adaptation for prognostics in a particle filtering framework."* International Journal of Prognostics and Health Management Volume 2 (color) (2011): 61.

Saha, Bhaskar, et al. "Prognostics methods for battery health monitoring using a Bayesian framework." *Instrumentation and Measurement, IEEE Transactions on* 58.2 (2009): 291-296.

Saha, Bhaskar, Kai Goebel, and Jon Christophersen. "Comparison of prognostic algorithms for estimating remaining useful life of batteries." *Transactions of the Institute of Measurement and Control* (2009).

Saitoh, Y., Iida, K., Sawada S., Shimizu, K. and Hattori, "Dependency of Contact Resistance on Load," Proc. of IEEE Holm conf., pp.70-75, 2007

Salami, M.J.E., Gani, A., Pervez, T., "Machine condition monitoring and fault diagnosis using spectral analysis techniques", *Proceedings of the First International Conference on Mechatronics (ICOM '01)*, vol. 2, Kuala Lumpur, Malaysia, 2001, pp. 690–700.

Satish, B. and Sarma, N., "A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors," in *Proc IEEE Power Engin. Soc. General Meet.*, 2005, pp. 2291–2294.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., and Schwabacher, M., "Metrics for evaluating performance of prognostic techniques", *IEEE International Conference on Prognostics and Health Management*, 2008.

Saxena, A., Celaya, J., Saha, B., Saha, S., and Goebel, K., *"*On applying the prognostics performance metrics", Annual Conference of the PHM Society, 2009.

Saxena, A., Celaya, J., Saha, B.,and Goebel, K., *"*Evaluating algorithm performance metrics tailored for prognostics", *IEEE Aerospace Conference*, 2009.

Saxena, A., Celaya, J., Saha, B., and Goebel, K., *"*Evaluating algorithm performance metrics tailored for prognostics", IEEE Aerospace Conference, 2009.

Saxena, Abhinav, and Ashraf Saad. "Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems." *Applied Soft Computing* 7.1 (2007): 441-454.

Scho¨mig, A. and Rose, O. ''On the Suitability of the Weibull Distribution for the Approximation of Machine Failures,'' *Proceedings of the 2003 Industrial Engineering Research Conference,* Portland, OR, May 18–20, 2003.

Senesky, Debbie G., et al. "Harsh environment silicon carbide sensors for health and performance monitoring of aerospace systems: A review." *Sensors Journal, IEEE* 9.11 (2009): 1472-1478.
Sharda, Ramesh. "Neural networks for the MS/OR analyst: An application bibliography." *Interfaces* 24.2 (1994): 116-130.

Si, X., Wang, W., Hu, C. and Zhou, D. "Remaining useful life estimation - A review on the statistical data driven approaches", *European Journal of Operational Research*, vol. 213, no. 1, pp. 1-14.,(2011)

Sikorska, J. Z., Hodkiewicz, M. and Ma, L. (2011), "Prognostic modelling options for remaining useful life estimation by industry", *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1803-1836.

Simani, S., Fantuzzi, C. Patton, R.J., "Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques", Springer, London, 2003.
Simon, Dan. "*Optimal state estimation: Kalman, H infinity, and nonlinear approaches*", John Wiley & Sons, 2006.

Singleton, R. Strangas, E. and Aviyente, S. "Extended kalman filtering for remaining-useful-life estimation of bearings," *Industrial Electronics, IEEE Transactions on,* vol. 62, no. 3, pp. 1781–1790, March 2015.

Skormin, V.A., Popyack, L.J., Gorodetski, V.I., Araiza, M.L., Michel, J.D., "Applications of cluster analysis in diagnostics-related problems", *Proceedings of the 1999 IEEE Aerospace Conference*, vol. 3, Snowmass at Aspen, CO, USA, 1999, pp. 161–168.

Slade P.G., "Opening contacts: the transition from the molten metal bridge to the electric arc", *IEICE Trans. Electron.* E93-C (6): 1380–1386, 2010.

Slade P.G., "The transition to the metal phase arc after the rupture of the molten metal bridge for contacts opening in air and vacuum", *Proceedings of the 54th IEEE Holm Conference on Electrical Contacts:* 1–8, 2008.

Slade, P.G.,"*Electrical Contacts: Principles and applications",* CRC Press, 2013.

Slade, P.G., "The transition from the molten metal bridge to the metallic phase bridge column arc between electrical contacts opening in vacuum", *Proceedings of the 23rd International Symposiumon Discharges and Electrical Insulation in Vacuum,* 198–201, 2008.

Slade, P.G., "The Vacuum Interrupter, Theory, Design and Application*",* Boca Raton, FL: CRC Press, 2008.

Slade, P.G., Hoyaux, M.F., "The effect of electrode material on the initial expansion of an arc in vacuum", *IEEE Trans. Parts Hyb. Pack*. PHP–8: 35–47, 1972.

Slade,P.G., Nahemow, M.D., "Initial separation of electrical contacts carrying high currents", *J. Appl. Phys.* 42(9): 3290–3297, 1971.

Smythe, WR. "*Static and Dynamic Electricity"*, New York: McGraw-Hill, 1968.

Sohn, Hoon, Keith Worden, and Charles R. Farrar. "Statistical damage classification under changing environmental and operational conditions." *Journal of Intelligent Material Systems and Structures* 13.9 (2002): 561-574.

Stack, J.R., Habetler, T.G., Harley, R.G., "Bearing fault detection via autoregressive stator current modelling", *IEEE Transactions on Industry Applications* 40 (2004) 740–747.

Steinmetz, C. P., "Electric power into light, Section VI. The Arc," *Trans.Amer. Inst. Elect. Eng*., vol. 25, p. 802, 1906.

Stokes, A. D. and Oppenlander, W. T., "Electric arcs in open air", *J. Phys.D, Appl. Phys.,* vol. 24, no. 1, pp. 26–35, Jan. 1991.

Sun, Bo, et al. "Benefits analysis of prognostics in systems." *Prognostics and Health Management Conference, 2010. PHM'10,* IEEE, 2010.

Sutherland, H., RePoFf, T., House, M. and Flicking, G., "Prognostics, A new look at statistical life prediction for condition-based maintenance," in *Proc. IEEE Aerospace Conf.*, 2003, vol. 7, pp. 3131–3136.

Swanson, D. C., "A general prognostic tracking algorithm for predictive maintenance," in *Proc. IEEE Aerospace Conf.*, 2001, vol. 6, pp. 2971–2977.

Swingler, Jonathan, and Alan Sumption. "Arc erosion of AgSnO2 electrical contacts at different stages of a break operation." *Rare Metals* 29.3 (2010): 248-254.

Swingler, Jonathan, and John W. McBride. "Modeling of energy transport in arcing electrical contacts to determine mass loss." *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A* 21.1 (1998): 54-60.

Timsit, RS., "On the evaluation of contact temperature from potential-drop measurements", *IEEE Trans Comp Hyb Manuf Tech* CHMT-6: 115, 1983.

Tsay, R.S., "Time series and forecasting: brief history and future research", *Journal of the American Statistical Association 95* (450) (2000) 638–643.

Turner, HW, Turner, C, "Discontinuous contact erosion", *Proc 3rd Int'l Research Symp on Elec Cont Phen* 309–320, 1966.

Turner, HW, Turner, C, "The erosion of heavy current contacts and material transfer produced by arcing", *Proc 4th Int'l Res Symp on Elec Cont Phen* 196–200, 1968.

Uckun, Serdar, Kai Goebel, and Peter JF Lucas. "Standardizing research methods for prognostics." *Prognostics and Health Management, 2008. PHM 2008. International Conference on*. IEEE, 2008.

Utsumi, S., Kawasaki, Z., Matsu-Ura, K., Kawada, M., "Use of wavelet transform and fuzzy system theory to distinguish wear particles in lubricating oil for bearing diagnosis", *Electrical Engineering in Japan 134* (2001) 36–44.

Utsumi, T., "Theoretical and experimental investigations of the dynamic molten bridge", *IEEE Trans. Parts Mater. Pack*. PMP 5(1): 62–68, 1969.

Vachtsevanos, Lewis, Roemer Hess and Wu. 2006. "Intelligent Fault Diagnosis and Prognosis for Engineering Systems", New Jersey: John Wiley & Sons.

Van, A. R., and Warrington, C., "Reactance relays negligibly affected by arc impedance," *Elect. World, vol. 98*, no. 12, pp. 502–505, Sep. 1931.

Vaseghi, Saeed V. "*Advanced digital signal processing and noise reduction*", John Wiley & Sons, 2008.

Voss, Richard F.; Clarke, John (1976-01-15). "Flicker (1/*f*) noise: Equilibrium temperature and resistance fluctuations". *Physical Review B*. **13** (2): 556–573.

Wakatsuki, N., Homma, H., “Melting phenomena and arc ignition of breaking relay contacts”, *Proceedings of the 54th IEEE Holm Conference on Electrical Contacts:* 15–20, 2008.

Wang, W.J., McFadden, P.D., “Early detection of gear failure by vibration analysis II. Interpretation of the time–frequency distribution using image processing techniques”, *Mechanical Systems and Signal Processing 7* (1993) 205–215.

Wang, Wilson Q., M. Farid Golnaraghi, and Fathy Ismail. "Prognosis of machine health condition using neuro-fuzzy systems." *Mechanical Systems and Signal Processing* 18.4 (2004): 813-831.

Williams, Z. Benefits of IVHM: an analytical approach. In Proceedings of the 2006 IEEE Aerospace Conference, Big Sky, Montana, USA, 4–11March 2006, paper no. 1507.

Wu, W., Hu, J.,Zhang, J., “Prognostics of machine health condition using an improved ARIMA-based prediction method”, *IEEE*, Harbin, China, 2007, pp. 1062–1067.

Xiong, Y., Cheng, X., Shen, Z. J., Mi, C., Wu, H. and Garg, V. K. (2008), "Prognostic and warning system for power-electronic modules in electric, hybrid electric, and fuel-cell vehicles"*, IEEE Transactions on Industrial Electronics*, vol. 55, no. 6, pp. 2268-2276.

Xu, J. and Xu, L., “Health management based on fusion prognostics for avionics systems,” *J. Syst. Engin. Electron.*, vol. 22, no. 3, pp. 428–436, 2011.

Xuerong, Ye, et al. "Degradation failure model of electromagnetic relay." *Electrical Contacts (ICEC 2012), 26th International Conference on*. IET, 2012.

Xuerong, Ye, Yu Qiong, and Zhai Guofu. "Reliability assessment for electromagnetic relay based on time parameters degradation." *Electronic Packaging Technology & High Density Packaging (ICEPT-HDP), 2010 11th International Conference on*. IEEE, 2010.

Xuerong, Ye, Yu Qiong, and Zhai Guofu. "Reliability assessment for electromagnetic relay based on time parameters degradation." *Electronic Packaging Technology & High Density Packaging (ICEPT-HDP), 2010 11th International Conference on*. IEEE, 2010.

Yan,J. and Lee, J., “A hybrid method for on-line performance assessment and life prediction in drilling operations,” in *Proc. IEEE Int. Conf. Automat. Logistics*, 2007, pp. 2500–2505.

Yao, Fang, et al. "Research on the failure diagnostics parameters and the reliability prediction model of the electrical contacts." *Electrical Contacts-2006. Proceedings of the 52nd IEEE Holm Conference on Electrical Contacts*. IEEE, 2006.

Yao, Fang, et al. "Research on uncertainty of bounce time for electromagnetic relay and its application in operating reliability estimation." *Electrical Contacts (ICEC 2012), 26th International Conference on*. IET, 2012.

Yao, Fang, et al. "Research on uncertainty of bounce time for electromagnetic relay and its application in operating reliability estimation." *Electrical Contacts (ICEC 2012), 26th International Conference on*. IET, 2012.

Yin, C. Y., et al. "A physics-of-failure based prognostic method for power modules." *Electronics Packaging Technology Conference, 2008. EPTC 2008. 10th*. IEEE, 2008.

Yu, Q., Qi M., Wang, S., Zhai, G., "Research on Life Prediction Based on Wavelet, Transform and ARMA Model for Space Relay", *4th IEEE Conference on Industrial Electronics and Applications*, (2012).

Yu, Qiong, et al. "Research on life prediction based on wavelet transform and ARMA model for space relay." *2009 4th IEEE Conference on Industrial Electronics and Applications*. IEEE, 2009.

Zhai, Guofu, Weiwei Fan, and Weiyan Liang. "Study on contact resistance dynamic characteristics of space electromagnetic relay." *Computational Engineering in Systems Applications, IMACS Multiconference on*. Vol. 1. IEEE, 2006.

Zhang, Xian, and Pierluigi Pisu. "An unscented kalman filter based approach for the health-monitoring and prognostics of a polymer electrolyte membrane fuel cell." *a a* 1, 2012

Zhang, Xiaodong, et al. "An integrated approach to bearing fault diagnostics and prognostics." *American Control Conference, 2005. Proceedings of the 2005*. IEEE, 2005.

Zhao, Fagang, et al. "Neuro-fuzzy based condition prediction of bearing health." *Journal of Vibration and Control* (2009).

Zhao, Fuqiong, et al. "An Integrated Prognostics Method Under Time-Varying Operating Conditions.", *IEEE Transactions on Reliability,* 64.2 (2015): 673-686.

Zhao, Fuqiong, Zhigang Tian, and Yong Zeng. "Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method." *Reliability, IEEE Transactions on* 62.1 (2013): 146-159.

Zhao, K. Ling, F., Levari, H. and Proakis, J. "Sliding window order-recursive least-squares algorithms", *IEEE Trans. Signal Processing, vol. 42*, pp.1961 -1972 1994.

Zhu, Kunpeng, Yoke San Wong, and Geok Soon Hong. "Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results." *International Journal of Machine Tools and Manufacture* 49.7 (2009): 537-553.

Zio E, Maio FD. "*A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system."* Reliability Engineering and System Safety 2010; 95: 49–57.

Zio E, Peloni G. "*Particle filtering prognostic estimation of the remaining useful life of nonlinear components."* Reliability Engineering and System Safety 2011; 96(3):403–9.

Zio E, Peloni G., "Particle filtering prognostic estimation of the remaining useful life of nonlinear components", *Reliability Engineering and System Safety*, 2011.