Intelligent Prognostics of Machinery Health
Utilising Suspended Condition Monitoring Data

Aiwina Soong Yin Heng
Bachelor of Engineering (Mechanical)
Queensland University of Technology, Australia

Thesis submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy

School of Engineering Systems
Faculty of Built Environment and Engineering
Queensland University of Technology
2009
The ability to forecast machinery failure is vital to reducing maintenance costs, operation downtime and safety hazards. Recent advances in condition monitoring technologies have given rise to a number of prognostic models for forecasting machinery health based on condition data. Although these models have aided the advancement of the discipline, they have made only a limited contribution to developing an effective machinery health prognostic system. The literature review indicates that there is not yet a prognostic model that directly models and fully utilises suspended condition histories (which are very common in practice since organisations rarely allow their assets to run to failure); that effectively integrates population characteristics into prognostics for longer-range prediction in a probabilistic sense; which deduces the non-linear relationship between measured condition data and actual asset health; and which involves minimal assumptions and requirements.

This work presents a novel approach to addressing the above-mentioned challenges. The proposed model consists of a feed-forward neural network, the training targets of which are asset survival probabilities estimated using a variation of the Kaplan-Meier estimator and a degradation-based failure probability density estimator. The adapted Kaplan-Meier estimator is able to model the actual survival status of individual failed units and estimate the survival probability of individual suspended units. The degradation-based failure probability density estimator, on the other hand, extracts
population characteristics and computes conditional reliability from available condition histories instead of from reliability data. The estimated survival probability and the relevant condition histories are respectively presented as “training target” and “training input” to the neural network. The trained network is capable of estimating the future survival curve of a unit when a series of condition indices are inputted.

Although the concept proposed may be applied to the prognosis of various machine components, rolling element bearings were chosen as the research object because rolling element bearing failure is one of the foremost causes of machinery breakdowns. Computer simulated and industry case study data were used to compare the prognostic performance of the proposed model and four control models, namely: two feed-forward neural networks with the same training function and structure as the proposed model, but neglected suspended histories; a time series prediction recurrent neural network; and a traditional Weibull distribution model. The results support the assertion that the proposed model performs better than the other four models and that it produces adaptive prediction outputs with useful representation of survival probabilities.

This work presents a compelling concept for non-parametric data-driven prognosis, and for utilising available asset condition information more fully and accurately. It demonstrates that machinery health can indeed be forecasted. The proposed prognostic technique, together with ongoing advances in sensors and data-fusion techniques, and increasingly comprehensive databases of asset condition data, holds the promise for increased asset availability, maintenance cost effectiveness, operational safety and – ultimately – organisation competitiveness.
KEYWORDS

Artificial neural networks; condition-based maintenance; condition monitoring; prognostics; reliability; suspended data.
TABLE OF CONTENTS

ABSTRACT ......................................................................................................................... I
KEYWORDS .................................................................................................................... III
TABLE OF CONTENTS .................................................................................................... IV
LIST OF TABLES ............................................................................................................... VII
LIST OF FIGURES ........................................................................................................... VIII
NOMENCLATURE ........................................................................................................... XI
STATEMENT OF ORIGINALITY ................................................................................... XIV
ACKNOWLEDGMENTS .................................................................................................... XV
CHAPTER 1 INTRODUCTION ......................................................................................... 1
  1.1 PROBLEM STATEMENT ....................................................................................... 4
  1.2 SCOPE OF RESEARCH ....................................................................................... 5
  1.3 ORIGINALITY AND CONTRIBUTION ................................................................. 6
  1.4 STRUCTURE OF THE THESIS ........................................................................... 9

CHAPTER 2 BACKGROUND AND PRELIMINARY LITERATURE REVIEW .................. 12
  2.1 MACHINERY MAINTENANCE .............................................................................. 12
    2.1.1 Key Elements of Effective Condition-Based Maintenance ............................... 15
    2.1.2 Data Acquisition ............................................................................................ 15
    2.1.3 Data Processing ............................................................................................. 17
    2.1.4 Maintenance Decision Support: Diagnostics and Prognostics ....................... 21
  2.2 SUMMARY ........................................................................................................... 22
CHAPTER 3 MACHINERY FAILURE AND CURRENT PROGNOSTIC TECHNIQUES ................................................................. 23

3.1 METHODS FOR PREDICTING MACHINERY FAILURE ................................................................................................. 23
  3.1.1 Prognostic Approaches – Condition Based Prediction ......................................................................................... 28
  3.1.2 Integrated Approaches – Prediction Based on Both Reliability and Condition Data .............................................. 40

3.2 REMAINING CHALLENGES IN ROTATING MACHINERY PROGNOSTICS ................................................................. 44
  3.2.1 Use of Censored Event Data and Censored CM Data ............................................................................................. 44
  3.2.2 Integration of Reliability and Prognostics .................................................................................................................. 46
  3.2.3 Non-linear Relationship between CM Indices and Actual Health of Monitoring Unit ......................................................... 47
  3.2.4 Accuracy of Assumptions and Practicability of Requirements ............................................................................. 48

3.3 SUMMARY ........................................................................................................................................................................ 50

CHAPTER 4 DESIGN OF A NOVEL PROGNOSTIC MODEL ................................................................................................. 52

4.1 ENABLING TECHNIQUES FOR THE DEVELOPMENT OF PROPOSED MODEL .......................................................... 53
  4.1.1 Artificial Neural Networks ........................................................................................................................................ 53
  4.1.2 Statistical Analysis of Survival/Failure Time ............................................................................................................. 65
  4.1.3 Kernel Probability Density Estimation ....................................................................................................................... 77

4.2 ARCHITECTURE OF THE PROPOSED PROGNOSTIC MODEL ................................................................................... 82
  4.2.1 Kaplan-Meier Estimation of Survival Probability ..................................................................................................... 84
  4.2.2 Failure PDF Estimation Based on Degradation Data ................................................................................................. 87
  4.2.3 Final Target Outputs for ANN Training ...................................................................................................................... 96
  4.2.4 The Proposed FFNN as a Hybrid of Sequence Reproduction Networks and Pattern Classification Networks ...................................................................................................................... 97

4.3 SUMMARY ........................................................................................................................................................................ 100

CHAPTER 5 MODEL VALIDATION USING COMPUTER SIMULATED DATA ............................................................................... 104

5.1 A MODEL FOR THE SIMULATION OF BEARING DEGRADATION HISTORIES .............................................................. 105
  5.1.1 Modelling Defective Bearing Vibration Signatures ................................................................................................. 105
  5.1.2 Simulation of Bearing Failure Histories ..................................................................................................................... 112

5.2 PROCEDURE OF MODEL VALIDATION USING SIMULATED SIGNALS ........................................................................ 117
  5.2.1 Processing of Simulated Signals ............................................................................................................................... 118
  5.2.2 Design of Performance Evaluation .......................................................................................................................... 124
  5.2.3 Determining the Structure of the Prognostic Models ............................................................................................... 128

5.3 SIMULATION RESULTS AND DISCUSSION .................................................................................................................... 137
  5.3.1 Results of Assessment I Produced by the Proposed Model .................................................................................... 138
  5.3.2 Results of Assessment I Produced by Model A ......................................................................................................... 142
  5.3.3 Results of Assessment I Produced by Model B ......................................................................................................... 145
  5.3.4 Results of Assessment I Produced by Model C ......................................................................................................... 147
  5.3.5 Results of Assessment I Produced by Model D ......................................................................................................... 148
  5.3.6 Overall Model Comparison for All Assessments ........................................................................................................ 149

5.4 SUMMARY ........................................................................................................................................................................ 150

CHAPTER 6 MODEL VALIDATION THROUGH INDUSTRY CASE STUDY ........................................................................... 152
6.1 Pump Condition Monitoring at Irving Pulp and Paper Mill.................152
6.2 The Data ........................................................................................................154
6.3 Prognostic Modelling ..................................................................................157
6.4 Assessment Results ....................................................................................162
  6.4.1 Results of Assessment I by the Proposed Model.................................163
  6.4.2 Results of Assessment I by Model A ....................................................168
  6.4.3 Results of Assessment I by Model B ....................................................171
  6.4.4 Results of Assessment I by Model C ....................................................174
  6.4.5 Results of Assessment I by Model D ....................................................175
  6.4.6 Overall Model Comparison for All Assessments...............................177
6.5 Summary ........................................................................................................178

CHAPTER 7 CONCLUSION AND FUTURE WORK...........................................180
  7.1 Summary ......................................................................................................180
  7.2 Future Work ...............................................................................................183
  7.3 Final Remarks .............................................................................................185

REFERENCES ......................................................................................................186
LIST OF TABLES

Table 2-1: Comparison of conventional maintenance approaches and the CBM approach ................................................................. 14
Table 3-1: List of rotating machinery health prediction methods and their merits and limitations ............................................................................................................. 24
Table 5-1: Datasets simulated for model training and testing ............................................. 118
Table 5-2: Datasets used for training in the simulation assessments ............................... 125
Table 5-3: Determining the number of hidden layers for FFNNs [159] ............................. 129
Table 5-4: Prediction output of the proposed model for simulated dataset No. 35 in Assessment I ........................................................................................................... 140
Table 5-5: Predicted survival probabilities produced by Model A for the test-set No.1 in Assessment I .......................................................... 144
Table 5-6: Predicted survival probabilities produced by Model B for dataset No.35 in Assessment I .......................................................... 146
Table 5-7: Penalty for the five models in each assessment ............................................. 149
Table 6-1: Pump histories used for model training and testing ..................................... 158
Table 6-2: Histories used for training in Assessments I, II and III ................................. 162
Table 6-3: Survival probabilities predicted by the proposed model for the 1st test-set (History No.30) in Assessment I .............................................. 165
Table 6-4: Predicted survival probabilities produced by Model A for the 1st test-set (History No.30) in Assessment I .............................................. 169
Table 6-5: Predicted survival probabilities produced by Model B for the 1st test-set (History No.30) in Assessment I .............................................. 172
Table 6-6: Penalty for the five models in each assessment ............................................. 177
## LIST OF FIGURES

| Figure 1-1: Plant maintenance expenditure in the United States | 3 |
| Figure 3-1: The adaptive prognostic model for rolling element bearings proposed by Li et al. [63] | 29 |
| Figure 3-2: Combination of a stochastic version of Yu-Harris bearing life equations with condition data for remaining useful life prediction [69] | 31 |
| Figure 3-3: Architecture of the Neuro-fuzzy system used for forecasting gear monitoring indices by Wang et al. [79] | 35 |
| Figure 3-4: Prediction of a bearing’s failure time using exponential parameters [82] | 37 |
| Figure 3-5: The inverted bearing dynamics model for recovering the defect impulse proposed by Li and Shin [88] | 39 |
| Figure 3-6: The relationship between PCM and PHM defined by Sun et al. [94] | 42 |
| Figure 3-7: Timeline of the operational life of a machine component | 45 |
| Figure 3-8: Assumption that failure takes place at the instant of time when the condition parameter exceeds a predefined threshold | 48 |
| Figure 4-1: Feed-forward neural network | 59 |
| Figure 4-2: Recurrent neural network | 60 |
| Figure 4-3: Radial basis function network | 61 |
| Figure 4-4: Self-organising map | 62 |
| Figure 4-5: Complete data | 66 |
| Figure 4-6: Right-Censoring / Suspension | 67 |
| Figure 4-7: Interval-Censoring | 68 |
| Figure 4-8: Left-Censoring | 68 |
| Figure 4-9: Weibull probability density function with different shape parameters | 70 |
| Figure 4-10: A sample Kaplan-Meier plot | 75 |
| Figure 4-11: A sample histogram | 78 |
| Figure 4-12: Six Gaussians (broken lines) and their sum (solid lines) [143] | 79 |
| Figure 4-13: Architecture of the FFNN used in the proposed prognostic model | 83 |
| Figure 4-14: Conventional failure PDF estimated from historical failure events | 88 |
| Figure 4-15: Instantaneous reliability based on historical degradation processes | 89 |
| Figure 4-16: Interpolation of data points for estimating the probability density in each operating interval | 92 |
Figure 4-17: Integral of the PDF between the observed degradation value of unit \( i \) and the threshold value ................................................................. 95
Figure 4-18: Integral of the PDF over all probable values equal to or higher than the observed degradation value of unit \( i \) ................................................................. 95
Figure 4-19: Sequence reproduction FFNN .................................................. 98
Figure 4-20: Pattern classification FFNN .......................................................... 98
Figure 4-21: Proposed FFNN as a hybrid of sequence reproduction networks and pattern classification networks ................................................................. 100
Figure 4-22: Architecture of the proposed prognostic model ........................................... 102
Figure 4-23: Utilisation of data by the proposed prognostic model ........................................... 103
Figure 5-1: Idealised vibration signature due to a localised defect in a bearing outer race [150] ................................................................. 107
Figure 5-2: Load distribution in a radially loaded bearing .................................................. 109
Figure 5-3: Simulated signal of a bearing with an inner race defect ........................................... 110
Figure 5-4: Simulated signal of a bearing with an inner race defect after being corrupted by noise ................................................................. 111
Figure 5-5: Signal of a bearing with multiple faults (a) Time domain plot of raw signal (b) Frequency spectrum of raw signal (c) Fault detection after ........................................... 112
Figure 5-6: Evolution of defect impulse in dataset No.1 – exponential growth with random fluctuations ................................................................. 114
Figure 5-7: Defect severity increasing exponentially (only one of every 10 datasets generated are plotted here) ................................................................. 115
Figure 5-8: Sample features extracted from simulated data ................................................................. 116
Figure 5-9: Examples of the computer-simulated degradation datasets ........................................... 117
Figure 5-10: Organisation of the weights of a 3-dimensional SOM before and after training ................................................................. 120
Figure 5-11: A sample trajectory of a 3-dimensional bearing degradation dataset in a SOM ................................................................. 121
Figure 5-12: Trending defect severity by using SOM Quantisation Error as a degradation indicator ................................................................. 122
Figure 5-13: The high-dimensional feature matrix was reduced to a one-dimensional input vector and fed into the prognostics model ................................................................. 123
Figure 5-14: Selecting the number of hidden neurons with forward selection ................................................................. 131
Figure 5-15: Tangent sigmoid transfer function ................................................................. 132
Figure 5-16: Saturated linear transfer function ................................................................. 133
Figure 5-17: Pure linear transfer function ................................................................. 135
Figure 5-18: Post-training regression analysis for network structure selection ................................................................. 136
Figure 5-19: The standard KM Estimation of survival probability for the 30 training datasets ................................................................. 137
Figure 5-20: Degradation of simulated dataset No.35 ................................................................. 139
Figure 5-21: The predicted survival probability curve for Dataset No.35 at \( t=64 \) ................................................................. 142
Figure 5-22: The RNN prediction results of test-set No.1 in Assessment I ................................................................. 147
Figure 5-23: Weibull plot by Model D ................................................................. 148
Figure 6-1: Vibration histories of the 12 failed bearings ................................................................. 156
Figure 6-2: The standard KM estimation of survival probability for the training histories ................................................................. 159
Figure 6-3: Original and interpolated condition data of History No.10 ................................................................. 160
Figure 6-4: Original and interpolated condition data of History No.17 ................................................................. 161
Figure 6-5: Degradation index of pump bearing in the 1st test-set (History No.30) ................................................................. 164
Figure 6-6: Graphical representation of the prediction output by the proposed model at selected time steps for the 1st test-set (History No.30) in Assessment I ......................... 167
Figure 6-7: Graphical representation of the prediction output by Model A at selected time steps for the 1st test-set (History No.30) in Assessment I ................................. 170
Figure 6-8: Graphical representation of the prediction output by Model B at selected time steps for the 1st test-set (History No.30) in Assessment I ................................. 173
Figure 6-9: The RNN prediction results for the 1st test-set (History No.30) in Assessment I ............................................................................................................. 175
Figure 6-10: Weibull plot by Model D ............................................................... 176
NOMENCLATURE

Abbreviations

AI       Artificial Intelligence
ANN      Artificial Neural Network
BPFI     Ball Pass Frequency, Inner Race
BPFO     Ball Pass Frequency, Outer Race
CBM      Condition-Based Maintenance
CM       Condition Monitoring
FEA      Finite Element Analysis
FFNN     Feed-Forward Neural Network
HMM      Hidden Markov Model
The health index estimated based on the fault severity of informatively suspended unit $i$ at repair/replacement

Greek Letters

$\Delta$ Time interval

$\mu_i$ The health index estimated based on the fault severity of informatively suspended unit $i$ at repair/replacement

Roman Abbreviations

$b$ Bias of an artificial neural network

d Number of delays in an artificial neural network

$F$ Failure distribution

$h$ Number of time intervals in the prediction horizon

$L$ The time interval in which a historical unit was last observed to be still surviving

$m$ Number of historical units under study

$R$ Reliability

$S$ Survival probability
$S_{KM}$ Survival probability estimated using the adapted Kaplan-Meier method

$S_{PDF}$ Survival probability estimated using the probability density estimation method

$t$ Time

$T$ Failure time

$w$ Weights in an artificial neural network

$y_i$ Random process of condition change in historical unit $i$

$Y_i$ Condition value for unit $i$

$Y_{thresh}$ Condition value at the failure threshold

**Subscripts**

$i, j, l$ Element number
STATEMENT OF ORIGINALITY

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signed ........................................

Date 24 - APR - 2009
I’d like to thank my family in Malaysia for their unwavering love and support. I also owe a debt of gratitude to Angeline, C.C. and Rachel – for providing me a home and making my stint in Brisbane a wonderful journey.

Sincere thanks to my supervisors, Prof. Andy Tan and Prof. Joseph Mathew, whose encouragement and guidance made this thesis work possible. Thanks are also due to Dr. Eric Kim, Dr. Sheng Zhang and Mr. Leith Hitchcock for their advice and friendship.

I would like to give special recognition to Prof. Andrew Jardine, Dr. Dragan Banjevic and Dr. Neil Montgomery at C-MORE, for their hospitality during my stay in Toronto and for helping me to achieve this important milestone in my life.

I am grateful to Prof. Bo-Suk Yang and his many students at IML for hosting my visit to Pukyong National University and for instilling in me the Korean values of cleverness and diligence.

Cheers to my friends, for making this so much fun. 😊

Lastly, special thanks with love to William for being my soulmate and best friend – all I need to get through anything.
CHAPTER 1 INTRODUCTION

Operational safety, asset availability and maintenance cost effectiveness have a direct impact on the competitiveness of organisations and nations. Unforeseen breakdowns not only can cause expensive downtime, but also safety and environmental detriments that may lead to injuries or fatalities, as well as enormous legal expenses. Frequent and unscheduled breakdowns also hinder the implementation of management strategies such as Just-In-Time (JIT) and Material Resource Planning (MRP), which are vital in reducing wasteful operation activities and inventory [1]. Today’s complex and advanced machines demand highly sophisticated and costly maintenance strategies. Even well back in 1981, domestic plants in the United States spent more than $600 billion to maintain their critical plant systems. This figure has doubled within two decades [2]. Currently American companies spend around $2 trillion a year on maintenance; half of this held up in inventories and the other half lies predominately in labour where, in the majority of maintenance organisations, maintenance craftsmen spend as little as two hours a day doing actual hands-on work activities [3]. The trend is similar in many other countries including Australia [4]. An even more alarming fact is that one-third to one-half of this expenditure is wasted through ineffective maintenance (see Figure 1-1). Industry can no longer absorb this incredible level of inefficiency. Many organisations have moved operations to countries that provide more economical labour and lower operating costs. If we are to regain competitiveness, where do we begin?

The answer to this question has many facets that must be addressed; nonetheless, it starts with improving the maintenance and reliability functions which impact asset
utilisation. In recent years, Condition-Based Maintenance (CBM) has become increasingly popular throughout industry. CBM is a philosophy of maintaining engineering assets based on non-intrusive measurements of their condition as well as on maintenance logistics [5]. Various condition monitoring (CM) technologies, such as vibration monitoring, thermal monitoring, surface and internal defect detection, wear debris analysis and oil analysis have been developed. Most of the existing CM systems focus on the acquisition of relevant asset condition information and the detection/classification of asset faults based on the acquired data. These technologies enable an existing problem to be detected, diagnosed and corrected before breakdowns or other serious consequences occur. However, the related and yet more important question is how to utilise this asset health information for predicting the remaining asset lifetime, optimising maintenance schedules and ultimately maximising organisational efficiency. Enhanced application of CBM is thus through prognosis – the forecast of an asset’s remaining operational life, future condition, or risk to complete operation. This new approach promises to minimise production downtime and to spare inventory, maintenance costs and operational hazards.
Currently, a number of models have been proposed in the research area of machinery fault prognosis. The physics-based models attempt to combine system-specific mechanistic knowledge, defect growth formulas and CM data to predict the propagation of a fault, whereas the data-driven approaches derive models directly from the acquired CM data. After an extensive literature review, several limitations of the existing models have been identified. For example, many of the existing models do not incorporate population characteristics information and suspended CM data into prognosis. This research is aimed at developing new practical methods for addressing these limitations.

The rest of this chapter will define the research problem, the boundaries limiting the scope of the investigation and the contribution of this work. Finally, a brief overview of the thesis is presented.
1.1 Problem Statement

Most of the previous attempts to forecast machinery failures in practice depend, directly or indirectly, on the judgement of experienced personnel. This subjective judgement is not always reliable and often unsustainable due to mobility of the workforce. In recent years, a number of computerised prognostic models have been proposed to forecast equipment health. These existing models have advanced the development of machinery prognosis. Nevertheless, several aspects need to be further investigated before prognostic systems can be reliably applied in real-life situations. Firstly, suspended CM data of historical units need to be directly modelled and utilised. The term “suspended CM data” in this work refers to the condition trending data of historical units that did not undergo failure. They are very common in practice, since machines are rarely run to failure. Treating historical suspensions as failures in prognostic modelling would result in underestimation of remaining asset life. Excluding suspended CM histories from modelling, on the other hand, would worsen the problem of data scarcity. Therefore, a good prognostic model must be able to utilise suspended CM histories and not just CM histories that end with a failure. Secondly, both reliability information and CM data need to be effectively integrated to enable longer-range prognosis. While CM data are corroborative data that actually reflect the state of individual operating units, they do not replace reliability data that reflect population characteristics. Both population characteristics information and individual unit condition data should be used to enable longer-range prediction. Thirdly, CM data are commonly taken to indicate the health of a monitored unit. However, the measured condition indices do not always deterministically represent the actual health of the monitored asset. The challenge here is to deduce the non-linear relationship between an asset’s actual survival status and the measured CM indices. Lastly, the existing prognostic models mostly come with a large number of assumptions such as the underlying system physics and failure distribution. However, most machines are complex and fault propagation is probabilistic in nature. A small amount of model parameter difference can result in large prediction errors. This work presents research into new practical methods for addressing these challenges.
1.2 Scope of Research

Even though the concept proposed in this research may be applied to various engineering assets, this work looks at rotating machine components and rolling element bearings in particular. Rotating machinery is common and critical in many organisations, and bearing failure is often the cause of breakdowns in this class of equipment. Other mechanical components failures include gear, shaft and seal failures.

This research studies the prognosis element of CBM. Fault diagnostics, which includes the detection and classification of a fault, has been extensively researched compared to prognostics; therefore, development of a diagnostics model is out of the scope of this work. Prognostics is considerably less well-understood than diagnostics and more difficult to formulate since its accuracy is subject to stochastic processes that have not yet occurred.

The prognostic model proposed in this work integrates statistical and artificial intelligence (AI) methods. Instantaneous reliability of historical items is first calculated using a variation of the Kaplan-Meier (KM) estimator and a degradation-based failure Probability density function (PDF). The estimated reliability will then be used as the training target for a Feed-Forward Neural Network (FFNN). The trained FFNN is able to estimate the future survival probability of a monitored item when the corresponding CM data are inputted. This technique models suspended histories, extracts population characteristics from CM histories, and ultimately trains an intelligent agent to estimate reliability based on the given CM data.

It is acknowledged that many different definitions of “failure” and suspension” exist. In this work, the definition of “failure” can vary according to the situation in which the proposed model is applied. Generally, an item is considered to have failed if it has ceased to perform the required function satisfactorily. Maintenance personnel usually define failure thresholds based on past experiences. These definitions, though general, suffice from the viewpoint of practical maintenance.
“Suspensions”, on the other hand, are defined as cases for which the failure has not yet happened, or may never happen. Suspensions considered in this work include true suspensions which are not due to deterioration factors (e.g. items opportunistically overhauled/replaced at the point of operation stoppages caused by the failure of other components, or items still remaining in operation at the end of data collection); and informative suspensions (items repaired/replaced to prevent failures because a fault has been detected).

To validate the proposed prognostic model, computer simulated data will be used for model training and testing. As the primary research goal is the development of a practical prognostic model, the model will also be validated through an industrial case study using in-situ field data.

1.3 Originality and Contribution

(a) Proposed prognostic model

The proposed prognostics approach will provide more robust prediction performance than the existing models. This work presents a novel approach of:

1. Including suspended CM histories in model training examples. Suspended data are fully utilised and amalgamated directly into prognostic modelling, not by using an artificial cut-off time, but rather by using the information that the item has survived up to the last observed time and the probability of the population surviving the following intervals. The inclusion of suspended CM histories in prognostic modelling avoids underestimation of remaining asset life and mitigates data scarcity problems.

2. Integrating all reliability data into prognostic modelling; and extracting population characteristics and conditional reliability from CM histories instead of from reliability event data. These procedures not only effectively
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

integrate population reliability and unit CM information for longer-range prediction, but also enable reliability to be calculated based on the true degradation processes of historical assets instead of on the historical time-to-failure data only.

3. Using an ANN to recognise the non-linear relationship between actual asset health and measured condition data. The model is able to learn to recognise how unit degradation is veiled in the non-deterministic changes in CM measurements and disregard fluctuations caused by non-deterioration factors.

4. Using a non-parametric approach to performing prognosis based on the true data obtained. The ability to predict without relying on assumptions such as those on system physics properties, degradation pattern, failure distribution and failure threshold avoids the potentially large errors brought about by making incorrect assumptions about those properties.

Outcome of this work includes the publication of three conference papers and a journal paper:


- Aiwina Heng, Andy C. C. Tan, and Joseph Mathew, *A general condition-based prognostics model for estimating the survival probability of


(b) Model for simulating bearing degradation vibration data

A model for simulating bearing degradation data has also been developed. This approach provides researchers with much needed data to design and test their prognostic models before real life test data are available. This model has resulted in the publication of two conference papers:


(c) Critical review of machinery prognostics literature

Literature on current prognostic techniques for rotating machines has been critically reviewed. This review synthesises existing publications, identifies knowledge gaps,
which alert researchers to opportunities for key contributions to the field, and was published in the *Mechanical Systems and Signal Processing* journal.


(d) **Performance metrics for comparison of prognostic model**

A penalty function was developed for comparing the proposed prognostic model to the other models in the literature. It considers the mean prediction accuracy and the prediction horizon of a prognostic model. The proposed performance metrics were presented in the cited above.

(e) **Validation of prognostic models through industry case study**

Industry pump data were obtained and used for a case study in order to validate the practicability of the proposed prognostic model.

### 1.4 Structure of the Thesis

This thesis is composed of 8 chapters. The following describes the subtopics contained in each chapter and how they are interlinked.

Chapter 1, introduces the topic and significance of the research and shows how the research objective has grown out of the unresolved problem identified in the research area. The scope of the research is defined so that the study is narrowed down to a specific and manageable research topic that is practical for the purposes of a doctoral thesis. The originality and contribution of this work are also presented.
Chapter 2 first presents the background information of machine maintenance and how it has evolved to its present state. Then, an overview of CBM is given. The overview introduces the three key elements in CBM: data acquisition, data processing and maintenance decision making (diagnostics and prognostics), and briefly reviews their development.

Having contextualised the research topic, a critical review of the existing prognostic methods is presented in Chapter 3. Section 3.1 first reviews the existing machinery prognostic models. The challenges remaining in the research area are then identified in the Section 3.2.

After thus identifying the limitations of the existing machine fault prognostic models, this study sets out to establish a model that addresses the identified challenges and to test the assertion that such a model is more reliable and practical in performing prognosis. The following four chapters detail the research contribution:

Chapter 4 discusses the development of the prognostic model proposed by the candidate to address the unresolved issues identified in the literature. To ensure better understanding of the proposed prognostic model, background information on techniques and theories used is first given in Section 4.1. The architecture and training principles of the proposed prediction model are then detailed in Section 4.2.

In Chapter 5, the proposed model is validated using computer simulated data. The first section of the chapter describes how vibration signatures of defective rolling element bearings are modelled and then turned into progressive degradation data. Having these simulated bearing degradation data, the model is tested and the procedure is described in the second section. Lastly, results are presented and analysed in the third section.

Since one of the main objectives of this work was to develop an innovative yet practicable prognostic model, the proposed model is evaluated through an industry
case study, which is presented in Chapter 6. The pump vibration histories obtained from Irving Pulp and Paper were used for the training, testing and comparison of the proposed model as well as the other three models described in the previous chapter. Section 6.1 briefly introduces the CM activities at Irving Pulp and Paper mill. In Section 6.2, the centrifugal pump vibration data gathered from the mill are described. Section 6.3 details the procedures and challenges of applying the prognostic models to real life prognosis. The last section presents the prognostic results of each model.

Chapter 7 presents the conclusion of the research and discusses the directions for future extension of this research.
CHAPTER 2  BACKGROUND AND PRELIMINARY LITERATURE REVIEW

This chapter gives the necessary background information for contextualising the research topic and is divided into two sections. Section 2.1 presents an introduction to machine maintenance, including its evolution and technologies. In Section 2.1.1, an overview of Condition-Based Maintenance (CBM) is given. The overview introduces the three key aspects in an effective CBM program: data acquisition, data processing and maintenance decision making (diagnostics and prognostics). Current research on each of these three aspects is summarised.

2.1 Machinery Maintenance

Machinery maintenance is intrinsically evolutionary. Maintenance strategies have progressed from run-to-failure maintenance, to preventive maintenance, then to condition-based maintenance managed by experts, and now towards a futuristic view of intelligent proactive maintenance.

Run-to-failure or corrective maintenance is the earliest form of maintenance, where maintenance actions only take place at breakdowns. To prevent catastrophic failures and emergency shutdowns, preventive maintenance was introduced in the 1950s. A
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

typical preventive maintenance scheme includes setting periodic interval for machine inspections and overhauls, regardless of the machine’s health condition. Bazovsky [6] pioneered the use of mathematical optimisation methods in preventive maintenance policies. Jardine [7] introduced decision models for determining optimal replacement or overhaul interval by analysing historical breakdown events and cost data. However, these fixed time maintenance policies were not well-received by most practitioners [8]. While these policies do sometimes reduce equipment failures, they are more labour-intensive, do not eliminate catastrophic failures and include unnecessary maintenance actions that may potentially result in incidental component damage. This is where CBM steps in. It is well known that 99% of failures are preceded by trendable indicators [9]. The best maintenance strategy is, therefore, to make maintenance decisions based on the current measurement of the equipment condition, thus avoiding unnecessary maintenance and facilitating timely maintenance when there is an indication of impending failure. CBM attempts to monitor equipment health based on these condition indicators without interrupting normal machine operations. A comparison of conventional maintenance approaches and the CBM approach is shown in Table 2-1.

Over the past few decades, technologies in data acquisition and fault diagnostics have become more mature. Information such as vibration signatures, acoustic emissions signatures and oil particle counts can be acquired, processed and analysed through state-of-the-art sensors, database software and parallel computation technologies. Nevertheless, new technologies often introduce new types of information that may not have been fully exploited. It has become a significant challenge for the research community to effectively integrate this new information into maintenance scheduling. The following section discusses three main aspects of CBM of mechanical equipments, with a focus on rotating machines.
### Table 2-1: Comparison of conventional maintenance approaches and the CBM approach

<table>
<thead>
<tr>
<th>Approach</th>
<th>Method</th>
<th>Problems</th>
</tr>
</thead>
</table>
| Corrective maintenance    | Run unit until it breaks down and then repair/replace it | • Failure can occur at unexpected time and location  
• Long production downtime  
• Failure of a unit can cause consequential damage to other units  
• Failure can cause safety hazards |
| Preventive maintenance    | Inspect/repair/replace unit at regular intervals | • Labour intensive  
• High inventory cost  
• Does not eliminate failures as units rarely fail at regular intervals  
• Regular disassembly of equipment and examination of every possible part can cause incidental damage |

<table>
<thead>
<tr>
<th>Approach</th>
<th>Method</th>
<th>Advantages</th>
</tr>
</thead>
</table>
| Condition-based Maintenance | Replace/repair when necessary based on the non-intrusive measurement of current unit condition | • Easy identification of the faulty component  
• Prevention of failures and increased asset availability  
• Reduced maintenance costs  
• Reduced inventory costs  
• Improved product quality  
• Improved safety |
2.1.1 Key Elements of Effective Condition-Based Maintenance

The three key elements of effective CBM are:

1. Data acquisition – collection and storage of machine health information

2. Data processing – conditioning and feature extraction/selection of acquired data

3. Decision making – recommendation of maintenance actions through diagnosis and/or prognosis

2.1.2 Data Acquisition

Machine run-time data are the basis of CBM programs. CBM data fall into two main categories: event data and CM data. Event data are records of occurrences such as installations, failures, restorations, preventive replacements or suspensions, as well as the occurrence times and dates. CM data for rotating machines, on the other hand, include vibration signatures, sound levels, temperature readings, oil particle counts and acoustic emission signatures. While CM data can be sampled, logged, formatted and stored through readily available data acquisition systems, event data are usually recorded manually by maintenance personnel.

The acquisition of event data is usually neglected and sometimes even regarded as not as important as CM data. Vlok [10] investigated the data collection practice in several companies and reported numerous shortcomings in data acquisition; for example, a component’s health state at the time of replacement and its operational age are seldom recorded. Collection of event data should not be overlooked because they assist in identifying the failure threshold for CM
data, determining whether a CM history ends in failure or suspension and in assessing the performance of condition indicators.

Acquisition of CM data are usually automated through computer software and hardware, such as the LabVIEW software and data acquisition cards produced by National Instruments. Software packages, such as the Computerised Maintenance Management System (CMMS), are also available now to help companies maintain a computer database of information about their maintenance operations. Various types of sensors have been developed, such as accelerometers, ultrasonic sensors and acoustic emission sensors. Vibration measurements through piezoelectric accelerometers are widely used for the CM of rotating machines because they offer good linearity, high durability and accuracy over a wide range of environmental conditions. Vachtsevanos et al. [11] identified four main requirements for optimal sensor placement for fault diagnosis: detectability, identifiability, fault detection reliability and a requirement associated with limited resources. Sensors must be validated to ensure that they are not subjected to fault conditions. Goldman [12] stated that all vibration readings should be taken with the sensor perpendicular to the surface of interest and that vibration signals containing high frequencies must be taken with the sensor tightly screwed or glued to the surface to correctly sense the frequencies. Temperature-sensing thermocouples are also widely used sensors in industry because defects such as cracks generate heat when subjected to mechanical vibrations. In recent years, research has been conducted to design and develop smart sensors to provide advantages such as: ease of installation, self-identification, self-diagnosis, reliability, time awareness for coordination with other nodes, some software functions and DSP, and standard control protocols and network interfaces [13]. Wireless technologies, such as Bluetooth, have also been developed to enable convenient, cost-effective data communication.
2.1.3 Data Processing

Good data processing is essential to reliable diagnosis and prognosis of machinery faults. Data acquired from rotating machines are usually value-type (e.g. oil analysis data and temperature) or waveform-type (e.g. vibration and acoustic emission). Special attention is given to waveform type data as they require more processing and many techniques have been developed for their analysis and interpretation. Raw data acquired from sensors are almost always processed before they are used for further analysis. Appropriate data features need to be calculated, selected and/or extracted for effective diagnostics or prognostics of the monitored equipment. This subsection discusses (i) data conditioning and (ii) feature extraction and selection.

(i) Data Conditioning

Raw waveform signals acquired from sensors can be very noisy, of low signal-to-noise ratio and biased. A basic technique for conditioning signals is low-pass filtering, since noise generally dominates the desired signals at high frequencies [11]. The chosen cut-off frequency should be at least twice as high as the highest useful signal frequency. To de-noise the acquired signals, Time Synchronous Averaging (TSA) was used in gear fault detection to average raw vibration signals over a number of evolutions to cancel out random noise [14, 15]. McFadden [16] demonstrated that the rejection of periodic noise of a known frequency can be optimised by the selection of an appropriate number of averages. Karimi et al. [17] applied TSA to rolling element bearing vibration signals. Another efficient and well-known algorithm – Adaptive Noise Cancellation (ANC) – was proposed by Widrow and Goodlin [18] by solving the Wiener-Hopf equations with a recursive scheme. The noise cancellation principle has been found useful for separating rolling element bearing signals from other vibration to assist in bearing fault diagnosis. Tan and Dawson [19] applied ANC to enhance the vibration signatures from rolling element bearings. ANC requires two sensors: one for measuring the noise-corrupted signal and one for measuring the reference noise. With the reference noise correlated to the
noise in the corrupted signal, ANC cancels out the noise component in the corrupted signal and leaves the desired signal. Adaptive Line Enhancer (ALE) [20] and Self-Adaptive Noise Cancellation (SANC) [21] were proposed to reduce the number of required sensors to one, by replacing the noise reference with a delayed version of the measured signal itself. However, SANC comes with several drawbacks in actual applications, such as the difficulty in parameter setting at adaptation phase. To address these issues, Antoni and Randall [22] proposed to directly estimate the noise cancellation frequency gain in the frequency domain. This frequency-domain algorithm was verified through application to the separation of bearing signals from gear signals in gearboxes, as the former are often masked by the latter. Tan et al. [23] applied Blind Deconvolution (equalisation) to recovering desired bearing signals without any prior knowledge of the receiving channel. It was demonstrated that this algorithm works well with periodic noise and that the equaliser functions as a notch filter in removing the noise at corresponding frequencies. Qiu et al. [24] proposed a wavelet filter based method to enhance weak signatures for the identification of incipient bearing faults.

(ii) Feature Extraction and Selection

For accurate diagnosis and prognosis, data must first be turned into information before knowledge can be acquired. To turn waveform data into information, fault condition indicators (features) are extracted and/or selected from the acquired signals. Reliable features generally have the following attributes [11]:

- Computationally inexpensive to measure
- Mathematically definable
- Explainable in physical terms
- Characterised by large interclass mean distance and small interclass variance
- Insensitive to extraneous variables
- Uncorrelated with other features

Features can be computed in the time domain, frequency domain or the time-frequency domain.
Root Mean Square (RMS) is the most commonly used time-domain feature. It is a measure of energy content in a signal. Another useful feature in representing fault condition is Kurtosis. It is the normalised fourth central moment and describes the relative spikiness and flatness of a distribution as compared to a normal distribution. More advanced methods of time-domain analysis involve fitting the signal to a parametric time series model and extracting relevant features based on the fitted model. Mechefske and Mathew [25] demonstrated that rolling element bearing vibration signals can be adequately modelled by an Auto-Regressive (AR) function. Poyhonen [26] applied the AR method to vibrations signals of an induction motor and the model coefficients were used as features. Baillie and Mathew [27] compared three techniques of AR modelling – back propagation neural network, radial basis function and traditional linear AR model – in the diagnosis of rolling element bearing faults. It was reported that the AR technique requires much shorter lengths of data than traditional pattern classification techniques. However, application of the AR model or its variant, the Auto-Regressive Moving Average (ARMA) model, is of limited applicability in practice due to the complexity in modelling. Qiu et al. [24] used a Self-Organising Map (SOM) to combine several time-domain features into a single bearing fault condition indicator. Zhang et al. [28] used Principal Component Analysis (PCA) to extract the features computed from both the time and frequency domains of pump vibration signals.

Time domain signals can be transformed to frequency domain by means of Discrete Fourier Transform (DFT). Frequency domain contains useful information for rotating machinery fault diagnosis and prognosis, as each component fault has its characteristic fault frequency. Power spectrum analysis is carried out in order to examine the component’s characteristic frequency closely and then extract features from the signals. Almeida et al. [29] proposed a frequency domain technique for evaluating the global vibration amplitudes of rolling element bearings. Envelope analysis has also been widely applied to rolling element bearing vibration signals [30-32]. When a fault develops in a rolling element bearing, the vibration becomes amplitude modulated due to
periodic changes in the forces. Envelope analysis aims to filter away the low frequency vibrations, including the fault symptoms, in order to extract the modulated periodic information from the more sensitive and pure envelope signal. Besides power spectrum, other useful spectra, such as Cepstrum [33] and Bispectrum [34], have also been developed for rolling bearing signal processing. Cepstrum analysis helps detect harmonics and sideband patterns in power spectrum. Bispectrum analysis can provide more diagnostic information than power spectrum for non-Gaussian signals. Mechefske and Mathew [35] built a parametric AR model for low speed rolling element bearing signals and then estimated power spectrum based on the fitted model.

Some machinery faults produce non-stationary signals. In these instances, frequency-domain analysis is not suitable and the energy of waveform signals needs to be represented in two dimensional functions of time and frequency. Short Time Fourier Transform (STFT) [36] and Wigner-Ville distribution (WVD) [37] have been widely applied to gear fault detection. STFT divides signals into segments with a sliding window, and performs Fourier transform on each of the windowed signals. However, STFT is not capable of providing high frequency and resolution at the same time. WVD, on the other hand, provides increased resolution relative to STFT but involves interference terms. Continuous Wavelet Transform (CWT) was proposed for bearing fault diagnosis and prognosis. CWT is faster than WVD and STFT in data processing. It gives good frequency and low time resolution for low-frequency components, and low frequency and high time resolution for high-frequency components. There is no interference in its decomposition. Rubini and Meneghetti [38] applied CWT to incipient fault detection of rolling element bearings. It was found that the high sensitivity of this wavelet technique to transient phenomena makes it suitable for the identification of the periodic shocks generated by spalls on bearing surface. CWT also performed better than envelope spectrum when the bearing defect was flattened by the contacting surfaces and when the fault signals were masked by machine resonance and noise. Qiu et al. [39] and Huang et al. [40] used CWT to process bearing vibration signals for fault prognosis.
2.1.4 Maintenance Decision Support: Diagnostics and Prognostics

The information extracted from the acquired data is analysed in diagnostic and/or prognostic systems so that appropriate maintenance actions can be recommended. Increased automation and mechanisation have made computerised diagnostic and prognostic systems a valuable tool for maintenance personnel in making maintenance decisions, and may even replace maintenance experts in due time. Today’s concept of machine diagnosis comprises the automated detection and classification of faults, whereas machine prognosis is the automated estimation of the timing and likelihood of a failure. Prognostics promises to significantly reduce expensive downtime, spares inventory, maintenance labour costs and hazardous conditions. However, prognostics is a relatively new research area and is therefore not as well-understood as the other areas of CBM, namely data acquisition, data processing and diagnostics. This research focuses on the prognosis of machinery faults after their detection and identification. Therefore, a comprehensive review of machine fault diagnostic techniques is out of the scope of this work. Only some basic diagnostic concepts will be briefly discussed. Literature on prognostics will be reviewed in more detail in the next chapter.

For fault detection, techniques such as Statistical Process Control (SPC) [41] and Statistical Change Detection (SCD) [42] have been used to detect the deviation of the monitored signal from a reference signal or model that represents the normal condition. As for fault identification, most diagnostics algorithms proposed in the literature are based on classifying or clustering signal features into different fault categories based on their similarity. Numerous different methods have been proposed for fault classification. They include the nearest neighbour algorithm [43], case-based reasoning [44], support vector machine [45, 46], fuzzy inference [47], hidden Markov models [48, 49] and artificial neural networks [34, 50-54]. Even though AI techniques have dominated the field of diagnosis, some model-based techniques have also been
proposed. Loparo et al. [55] used a mathematical model for the study of rub impact and detection of raceway faults in rolling element bearings.

As already mentioned, diagnostic techniques have been extensively researched and even implemented; thus, the focus of this research is on the much less well-understood prognostics. The following chapter presents a more detailed review of prognostic methods.

2.2 Summary

CBM is a philosophy of maintaining engineering assets when necessary, and is based on non-intrusive measurements of their condition. As opposed to the traditional run-to-failure maintenance and preventive maintenance, CBM enables easy identification of the component at fault; reduces failure occurrences, expensive downtime, inventory and labour costs; and may also improve product quality and enhance safety. This chapter presented the three key elements of effective CBM: data acquisition, data processing and decision making based on the acquired information.

Most of the existing CM systems and research focus are on the acquisition of asset condition data and the detection/classification of asset faults. However, the related and yet more important question is how to utilise this information to predict remaining asset lifetime for the optimisation of asset utilisation and maintenance scheduling. The following chapter summarises the current prognostic methods for rotating machinery and identifies the remaining challenges in the research area.
CHAPTER 3  MACHINERY FAILURE AND CURRENT PROGNOSTIC TECHNIQUES

Chapter 2 has provided a contextual background to CBM and this chapter will focus on the existing developments for rotating machinery prognostics. Rotating machinery has received considerable attention because of its common use across industries, critical operating regimes, frequent failure modes and the availability of its condition measurements. The number of publications on rotating machinery prognosis has been increasing steadily over the past few years. Subsection 3.1 reviews the existing techniques for rotating machinery prognostics, placing them in context and showing how they add to the accumulation of knowledge in the area. Remaining challenges in the research area are identified in Subsection 3.2.

3.1 Methods for Predicting Machinery Failure

The existing methods for predicting rotating machinery failures can be grouped into the following three main categories:

1. Traditional reliability approaches - event data based prediction

2. Prognostic approaches - condition data based prediction
3. Integrated approaches - prediction based on both event and condition data

Table 3-1 lists the existing models under these three approaches and summarises their merits and limitations.

<table>
<thead>
<tr>
<th>APPROACH</th>
<th>MERITS</th>
<th>LIMITATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reliability – uses event data, e.g. replacement/failure times of historical units</td>
<td>• Population characteristics information enable longer-range forecast • Do not require CM</td>
<td>• Only provide general, overall estimates for the entire population of identical units - not necessarily accurate for individual operating units</td>
</tr>
<tr>
<td>Traditional reliability models (e.g. Weibull, Poisson, Exponential and Log-Normal distributions) [6, 56-62]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Prognosis – uses CM data, e.g. vibration measurements of operating units</td>
<td>• Can be highly accurate if physics of models remain consistent across systems • Require less data than data-driven techniques</td>
<td>• Real-life system physics is often too stochastic and complex to model • Defect-specific</td>
</tr>
<tr>
<td>(i) Physics-based prognostic models</td>
<td>• Least-square scheme enables adaptation of model parameters to changes in condition</td>
<td>• Defect area size is assumed to be linearly correlated to vibration RMS level • Least-square scheme similar to single-step adaptation in time series prediction • Material constants to be determined empirically</td>
</tr>
<tr>
<td>- Paris’ law crack growth modelling [63, 64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Paris’ law crack growth modelling with FEA [65, 66]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Forman law crack growth</td>
<td>• FEA enables material stress calculation based on bearing geometry, defect size, load and speed</td>
<td>• Performance relies on the accuracy of crack size estimation based on vibration data • Computationally expensive</td>
</tr>
<tr>
<td></td>
<td>• Relates CM data and crack growth physics to life models</td>
<td>• Simplifying assumptions need to be examined</td>
</tr>
</tbody>
</table>
### Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

<table>
<thead>
<tr>
<th>Modelling [67]</th>
<th>Calculating time to spall initiation and the time from spall initiation to failure</th>
<th>Model parameters yet to be determined for complex conditions (e.g. in shaft loading zone and plastic zones)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Fatigue spall initiation and progression model [68-70]</td>
<td>Cumulative damage since installation is estimated with consideration of operating conditions</td>
<td>Various physics parameters need to be determined</td>
</tr>
<tr>
<td>- Contact Analysis for Bearing Prognostics [71]</td>
<td>FEA enables material stress calculation based on bearing geometry, defect size, load and speed</td>
<td>Various physics parameters need to be determined</td>
</tr>
<tr>
<td>- Stiffness based damage rule model [72]</td>
<td>Relates bearing component natural frequency and acceleration amplitude to the running time and failure time</td>
<td>Computationally expensive</td>
</tr>
<tr>
<td>(ii) Data-driven prognostic models</td>
<td>Do not require assumption or empirical estimation of physics parameters</td>
<td>Generally require a large amount of data to be accurate</td>
</tr>
<tr>
<td>- Simple trend projection models [73-75]</td>
<td>Ease of calculation</td>
<td>Rely on past degradation pattern and can lead to inaccurate forecasts in times of change</td>
</tr>
<tr>
<td>- Time series prediction using ANNs [76-81]</td>
<td>Fast in handling multivariate analysis</td>
<td>Assumes that condition indices deterministically represent actual asset health</td>
</tr>
<tr>
<td>- Exponential projection using ANN [82]</td>
<td>Estimates actual failure time instead of condition index at future time steps</td>
<td>Assumes that failure occurs once the condition index exceeds a predetermined threshold</td>
</tr>
<tr>
<td>- Data interpolation using ANN [83]</td>
<td>Longer prediction horizon</td>
<td>Short prediction horizon</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Assumes that all bearing degradation follow an exponential pattern</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requires training one ANN for each historical dataset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requires training a separate ANN for each historical data set</td>
</tr>
</tbody>
</table>
| Models combining reliability and prognostics | - Utilise available information more fully for increased accuracy  
- Longer-range prediction | - Require both event and condition data to be accurate  
- Assumes an underlying distribution |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IP and PF interval representation using Weibull distribution [89]</td>
<td>- Combines reliability and CM data to narrow down the time-to-failure window</td>
<td></td>
</tr>
</tbody>
</table>
- Accurate transition probability calculation requires a relatively large amount of CM histories  
- PHM assumes that hazard changes proportionately with covariates and the proportionality constant is the |
| EXAKT (combining Proportional Hazards Models, transition probability and a cost model) | - Combines economic considerations with failure prediction in aiding maintenance decision  
- Identifies the order of significance of trending data features |  
- Identifies the order of significance of trending data features  
- Accurate transition probability calculation requires a relatively large amount of CM histories  
- PHM assumes that hazard changes proportionately with covariates and the proportionality constant is the |
The first approach listed in Table 3-1 is the traditional reliability estimation. It is based on the distribution of event records of a population of identical units. Many parametric failure models – such as Poisson, Exponential, Weibull, and Log-Normal distributions – have been used to model machine reliability. The most popular among them is the Weibull distribution due to its ability to accommodate various types of behaviour, including infant mortality in the “bath-tub” curve.

Reliability analyses have been extensively studied and developed over the past few decades and numerous books and articles have been published [such as Refs. [6, 56-62]. These classical reliability approaches basically use historical time-to-failure data to estimate the population characteristics (such as mean time to failure and probability of reliable operation). However, these approaches only provide general overall estimates for the entire population of identical units. This type of estimations is useful to manufacturers that produce units in high volumes but are of little value to end users. For example, a maintenance engineer would be more interested in the ongoing reliability information of a particular piece of component
currently running in the machine, rather than in the mean time to failure of the whole population of such a component.

To estimate the current condition of an operating unit, a more “engineering” approach to reliability based on the actual change in unit health is necessary. Recent developments in CM technologies have enabled the collection of non-intrusive degradation measurements of a unit in operation. These CM data are a rich source of information in reliability evaluation of individual units. Consequently, the research community has started to develop prognostic models that estimate the future health of a monitored unit based on acquired CM data. This second category of failure prediction models are reviewed in Section 3.1.1. While CM data are corroborative data that reflect the current health of an operating item, they do not replace reliability data that reflect population characteristics. The last category comprises models that integrate reliability data into prognostics and is reviewed in Section 3.1.2.

3.1.1 Prognostic Approaches – Condition Based Prediction

As can be seen in the second section of Table 3-1, most of the existing prognostic models can be divided into two main categories: (i) physics-based models and (ii) data-based models.

(i) Physics-based prognostic models

Physics-based models typically involve building technically comprehensive mathematical models to describe the physics of the system and failure modes, such as crack propagation and spall growth. These models attempt to combine system-specific mechanistic knowledge, defect growth formulas and CM data to provide “knowledge-rich” prognosis output.
A common physics-based approach is crack growth modelling. Li et al. [63, 64] modelled rolling element bearing defect growth using a variation of Paris Law

\[
\dot{D} = \frac{dD}{dt} = C_0 (D)^n
\]  

(3-1)

which states that the rate of defect growth \( \dot{D} \) is related to the instantaneous defect area \( D \), assuming a constant operating condition. The parameters \( C_0 \) and \( n \) are material constants that need to be determined empirically. To make the prognostic model adaptive to the variation in defect growths, the predicted bearing defect size was compared to the actual defect size, and adaptation was performed using the recursive least-square scheme. Since the actual instantaneous defect area size is unavailable without interrupting the machine operation, it was assumed to be linearly correlated to the measured vibration RMS level. It was shown in numerical simulation tests and bearing life tests that parameter fine-tuning needs to be conducted because the physics-based model alone is unable to capture and adapt to the stochastic characteristics of defect growth. It was also shown that a small amount of parameter difference results in large prediction errors with the increase of bearing running cycles.

![Figure 3-1: The adaptive prognostic model for rolling element bearings proposed by Li et al. [63]](image-url)
Oppenheimer and Loparo [67] modelled rotor shaft crack growth using the Forman law of linear elastic fracture mechanics. CM information was integrated into the physics model through observer filters. However, since it is impractical to identify the instantaneous defect area size during operation, this crack growth model also relies on directly estimating defect area size from vibration data. The authors also stated that some issues – such as the determination of the motion and loads at motor ends from measured data, and the determination of stress intensity at crack locations near shaft discontinuities or under complex shaft loading – still need to be addressed. Simplifying assumptions also needed to be examined before the technique could be applied to specific machinery.

Qiu et al. [72] explored the use of damage mechanics for bearing prognosis by considering bearing system as a single-degree-of-freedom vibratory system. They proposed a stiffness-based prognostic model based on vibration data and damage mechanics. The failure natural frequency and the acceleration amplitude were related to the running time and failure time established from damage mechanics. Least-square parameter estimation algorithm was applied to continuously minimise the prediction error. Their model uses several parameters, such as stiffness constants for damaged parts, undamaged parts and the total system.

Orsagh et al. [68, 69] used a stochastic version of the Yu-Harris bearing life equation [98] to predict spall initiation and the Kotzalas-Harris progression model to estimate the time to failure. This system consisted of three main modules: sensed data, current bearing health and future bearing health modules (Figure 3-2). CM data, as well as loading induced by engine speed and manoeuvres, were first inputted into a rolling contact fatigue model, which calculated cumulative damage sustained by the bearing since its installation. The fatigue model parameters were derived from the data using physics-based and empirical models. If a spall did not seem to exist, time to spall initiation was estimated and inputted into a spall progression model, which estimated the time from spall initiation to failure. On the other hand, if a spall was diagnosed, the
time-to-initiation estimation was bypassed and the spall progression forecast was performed directly.

Kacprzynski et al. [70] enhanced the above system by proposing a framework for physics-based prognostics integrating material-level models, system-level data fusion algorithms and parameter tuning techniques. A spiral bevel pinion gear of a helicopter gearbox was used as a case study. It was stated that uncertainty in certain factors such as gear geometry, contact, load and material properties limits the reliability of prognostic systems. Information such as statistical variation of fatigue and fracture properties, as well as compressive stresses, has a dominant impact on the prediction accuracy of physics-based models.
Marble and Morton from Sentient Corporation [71] described a comprehensive experimental study of bearing spall progression. Sentient has created a program called Contact Analysis for Bearing Prognostics (CABPro). This program uses FEA to calculate material stress surrounding a spall based on the bearing geometry, spall size, rolling element load and speed, and then determines the cycles to failure based on damage mechanics principles. Evidence from a physical testing indicated that the propagation of spalls under high lambda (central lubrication film thickness divided by RMS surface roughness) conditions is less stochastic than was previously assumed.

A spur gear crack prognostic model that integrates a 2D Finite Element Analysis (FEA) failure model with an Artificial Neural Network (ANN) crack size estimation algorithm was proposed by Li and Choi [65]. The FEM failure model calculates stress and strain fields based on tooth load, geometry and material properties, and then predicts the direction and increment of crack growth using the Paris law. Like other physics based models, this failure model requires the knowledge of crack geometry, which is not practical in real life applications. Thus, an ANN was used to estimate the crack size based on measured vibration data. The model considers a tooth to be failed when the transverse crack size reaches a predetermined percentage of the tooth thickness. The number of load cycles for the crack to reach that assumed threshold can be derived from the Paris law. The remaining life, in terms of pinion revolutions, can then be estimated. The ANN in this crack FEA model was later replaced with an embedded modelling algorithm to identify gear meshing stiffness and estimate crack size from measured vibration data [66]. A gear dynamics model was also used to simulate the meshing dynamics and to determine the dynamic load on the cracked tooth. Three gears were run to failure in the laboratory and the proposed model was calibrated with one dataset, and tested with the other two datasets. The maximal prediction error (the difference between the actual and predicted remaining life) was reported to be less than 7%.
(ii) Data-driven prognostic models

Data-driven approaches attempt to derive models directly from routinely collected CM data, instead of building models based on comprehensive system physics and human expertise. They are built based on historical records and produce prediction outputs directly in terms of CM data.

The conventional data-driven methods include simple projection models, such as the exponential smoothing [73] and autoregressive models [74]. One major advantage of these techniques is the simplicity of their calculations, which can be carried out on a programmable calculator. However, most of these trend forecasting techniques assume that there is some underlying stability in the monitored system. They also rely on past patterns of degradation to project future degradation. This reliance could lead to inaccurate forecasts when the trend changes. Most of these models follow the changing pattern with a time lag of at least one observation. Cempel [75] introduced the Tribo-vibroacoustical (TVA) model which can estimate the time to failure of a machine as well as forecasting the vibration amplitude or condition. The model was compared with a constant trend parabolic model, an exponential trend model and an adaptive trending model in predicting a rolling bearing’s peak vibration acceleration. It was reported that none of the forecasting techniques was able to predict the sudden change in the life curve.

ANN is currently the most commonly used data-driven technique in the prognostics literature. An ANN consists of a layer of input nodes, one or more layers of hidden nodes, one layer of output nodes and connecting weights. The network learns the unknown function by adjusting its weights with repetitive observations of inputs and outputs. The most simple ANN-based machinery prognostic approach was time series prediction models. Tse and Atherton [76] approached machine health prognosis as time series prediction where a sequence of CM indices was used to forecast the successive monitoring index values at the next time step. A recurrent neural network (RNN) model was applied to the prediction of non-linear sunspot data and the trending of vibration data from
several industrial machines. Yam et al. [77] used an RNN to perform one-step-ahead prediction of CM data from a planetary gear train. It was demonstrated that RNN is capable of learning the trend of non-linear temporal data and is suitable for time series prediction.

Wang and Vachtsevanos [78] developed a recurrent wavelet neural network (RWNN) to predict rolling element bearing crack propagation. An artificial crack was seeded in the bearing and later manually enlarged at each vibration data collection, assuming that crack size increases uniformly with time. Network training was reported to be satisfactory when 100 data points were used, even though the training took several hours to complete. The network needed to be retrained after each time step to produce good results. The crack size was measured and fed back into the network to update the prediction error.

Wang et al. [79] used a neuro-fuzzy (NF) network to predict spur gear condition value one step ahead. The fuzzy interference structure is determined by experts, whereas the fuzzy membership functions are trained by the neural network. Figure 3-3 depicts the NF network architecture. The nodes in Layer 1 transmit input signals to Layer 2, in which each node acts as a membership function. Fuzzy rule firing strengths are calculated in Layer 3 and normalized in Layer 4. Layer 5 combines the input variables, and the predicted output is obtained in Layer 6. The NF system performed much better than RNN when there was sufficient training data. However, it could not predict well when the training dataset was small or when there were fast dynamic fluctuations, such as during the chipping of gear tooth surface material or just prior to gear failure.
An adaptive training technique was then proposed by Wang [80] to improve the NF model. A feedback link was added to each node in the 2nd layer of the original NF model. This adaptive NF model was compared with the traditional NF model in forecasting the propagation of an artificially induced crack on a gear. It was reported that both models were capable of capturing the system’s dynamic behaviours. However, the adaptive model was able to recapture the dynamic characteristics in a much shorter time. However, further work is needed to extend the prediction horizon from single step to multiple steps ahead.

Feed-forward neural network (FFNN) has also been used to perform single-step-ahead prediction of rolling element bearing condition data by Shao and Nezu [81]. Multiple-step-ahead predictions were also performed simply by feeding the predicted value back into the network input until the desired prediction horizon
was reached. The authors also proposed some rules to vary the data sampling period according to the change ratio of consecutive condition variables.

So far, all the ANN models mentioned above only produced estimates of the future asset condition indices. Gebraeel et al. [82] attempted to predict the actual bearing failure time instead of the future condition index. Their thrust bearing prognosis was based on the assumption that all bearing degradation signals possess an inherent exponential growth of the form $ae^{\beta t}$, where $\alpha$ and $\beta$ are the exponential parameters that give the best fit. They proposed two models. In the first one, an FFNN was trained for each of the 25 training datasets acquired. Once the input vector (consisting of the defective frequency amplitude and its first six harmonics) was presented to the network, the network would output (i) the bearing’s operational age at that instance, (ii) exponential parameters $\alpha$ and $\beta$, and (iii) the predicted failure time. The predicted operational age was compared with the actual operational age (it was assumed that knowledge of the bearing’s actual operational age at any point of prediction would always be available). Weights would be assigned to the 25 networks according to their accuracies in predicting the bearing’s operational age. These weights were multiplied with (i) the predicted failure time or (ii) the predicted exponential parameters $\alpha$ and $\beta$ from each network. Remaining useful life was then found by solving for the projected failure time at the predetermined failure threshold (Figure 3-4). The second model grouped the bearings into several clusters and used a Generalized Regression Neural Network to compute a single regression function for the degradation signals of the bearings from the same clusters. One FFNN was trained for each cluster. The rest of the selection and prediction procedure was performed in the same fashion as in the first model. The models that used a weighted average of exponential parameters produced the best results. Ninety-two percent of the failure time predictions were within 20% of the actual bearing life.
Huang et al. [83] built on the above method and used an FFNN to predict the failure time of single-row deep-groove ball bearings based on 100 interpolated points. It was stated that 85% of the failure time predictions were within 20% of the actual bearing lifetime.

Particle filtering has also been employed to provide non-linear projection in forecasting the growth of a crack on a turbine engine blade [84]. The current fault dimension was estimated based on the knowledge of the previous state of the process model. The a priori state estimate was then updated using new CM data. To extend this state estimation to multi-step-ahead prediction, a recursive integration process based on both Importance Sampling and Kernel probability density function approximation was applied to generate state predictions to the desired prediction horizon.

Jantunen [85] used high-order regression functions to mimic bearing fault development and also to save trending data in a compact form. Fuzzy classification limits were set to define eight classes of predicted conditions or
health states of a rolling element bearing. Simplified fuzzy logic was then used to classify the bearing health state. Indication of time to failure or probability of failure was not provided.

Zhang et al. [86] proposed a recursive Bayesian technique to calculate instantaneous failure probability based on the joint density function of different CM data features. This failure probability value was compared with the previous value and the conditional reliability was only updated when the failure probability value increased. This method enabled reliability analysis and prediction based on the degradation process of historical units, rather than on failure event data. The prediction accuracy of this model relied strongly on the correct determination of thresholds for the various trending features.

The use of Hidden Markov Models (HMMs) in bearing fault prognosis was investigated by Zhang et al. [87]. In an HMM, a system is modelled to be a stochastic process in which the subsequent states have no causal connection with previous states. In the work of Zhang et al., one HMM was trained to recognise one type of cone-and-cup bearing faults based on the corresponding vibration data. The HMMs were also trained to estimate the fault states, such as “normal”, “nick”, “scratches”, “more nicks” and “failure”. The similarity between current state and failure state was used as the bearing degradation index, which was then extrapolated to estimate the time of exceeding a predetermined failure threshold. HMMs do not represent temporal structure adequately since their state durations follow an exponential distribution. Dong and He [99] have thus proposed a segmental Hidden Semi-Markov Model (HSMM). Unlike HMMs, which generate a single observation for each state, HSMMs generate a segment of observations and estimate the durations from training data. However, the other possible limitations of HMMs – such as the difficulty in relating the defined health-state change point to the actual defect progression – remain, since it is often impractical to physically observe a defect in an operating unit.

Li and Shin [88] argued that it is often impractical to estimate bearing fault severity directly from vibration data because such data can be affected by many
factors other than defect severity; for example, by machine operating conditions and vibrations from nearby machine components. They proposed a system identification method to identify a bearing dynamics model, using measured bearing vibration (Figure 3-5). This bearing dynamics model was then inverted to recover the impact impulse behind a measured vibration signal. A modelling error compensation scheme was used to refine the recovered impulse. The estimated relative severity index was reasonably accurate only when the signal-to-noise ratio was high, i.e. when spall severity was already 85% and with relatively high running speed.

![Figure 3-5: The inverted bearing dynamics model for recovering the defect impulse proposed by Li and Shin [88]](image)

Lee [100] presented a scaleable prognostic platform – the Watchdog Agent toolbox – along with a systematic approach to developing and deploying advanced prognostic tools. This approach enabled researchers and engineers to select the most suitable algorithms based on multiple criteria. The prognostic modules employed in the toolbox include Auto-Regressive Moving Average (ARMA) prediction, Match Matrix, Elman Recurrent Neural Network and Fuzzy Logic.
Data-driven models are often the most available solution in many practical cases where it is easier to gather data than to build accurate system physics models.

3.1.2 Integrated Approaches – Prediction Based on Both Reliability and Condition Data

The previous section presented conventional reliability models as well as condition-based prognostic models made possible by the new CM technologies. While CM data are a rich source of information for fault/failure prediction, it should be noted that they do not render reliability data unnecessary. Several valuable models have considered integrating reliability data into prognostics.

Goode et al. [89] proposed a statistical method to predict the remaining useful life of pumps in a hot strip steel mill. Alarm limits were first determined using the Statistical Process Control (SPC) theory, with the assumption that healthy state data follow a normal distribution. SPC theory states that at least 99.73% of subsequent measurement values should fall within the stable state band for the pump condition to remain in the “Installation to Potential failure” (IP) interval. Otherwise, the pump condition is considered to have entered the “Potential failure to Functional failure” (PF) interval. During the IP interval, failure probability in a time interval is calculated using a cumulative density function for the PF Weibull distribution, without considering CM data. When in the PF interval, the time to failure is calculated considering vibration data, as defined by the following equation

\[
TTF = \left( \gamma_{PF} + \eta_{PF} \right) \times \left( 1 - \frac{\ln \left( X(t) - LL \right)}{\ln \left( AL - LL \right)} \right) \left( \frac{1}{\ln \left( UL - LL \right) / (AL - LL)} \right)
\]

where \( X(t) \) is the vibration measurement value, \( LL \) the lower limit, \( AL \) the alarm limit, \( UL \) the upper limit, \( \beta \) and \( \eta \) are respectively the shape and characteristic life parameters derived from the analysis of historical failures. It was stated that,
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

given sufficient failure lead time and accurate CM measurements, this model enabled a more systematic approach to assessing the risk of machine failure and was applicable to most practical cases. Goode et al. commented that the current literature on remaining life prediction focused predominantly on reliability based or mathematically complex models. Thus, they saw a need for a simple prediction model that is readily applicable to industrial situations.

Jardine et al. [90, 91] applied Proportional Hazards Model (PHM) to condition (oil) data and failure event data of aircraft and marine engines. PHMs assume that hazard changes proportionately with covariates (asset condition in this case) and that the proportionality constant is the same at all times. The use of PHM for incorporating monitoring information in reliability calculation started in the 1980s [101-104]. In PHM, the hazard of a system at time $t$ is modelled as

$$
\lambda(t) = h_0(t)\exp(\sum \gamma_i Z_i(t))
$$

where $h_0(t)$ is a deterministic baseline hazard function and $Z(t) = (Z_1(t), Z_2(t), \ldots)$ is a vector of time-dependent covariates such as engine oil analysis data [105]. Jardine et al. used Weibull distribution to model the baseline hazard function

$$
h_0(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1}
$$

where $\beta$ and $\eta$ are the shape parameter and scale parameter for the distribution respectively, and estimated using the maximum likelihood method. The covariates were assumed to follow a non-homogeneous Markov stochastic process [92, 93]. A software package called EXAKT [93] was developed to calculate the optimal maintenance or replacement time intervals based on failure history (both event and CM data) as well as cost data.
Sun et al. [94] pointed out that PHM models the change in system hazard as a response to the change in CM covariates. For most situations however, the opposite is true; that is, the change in CM covariates is a response to the change in system hazard. Thus, a Proportional Covariates Model (PCM) was proposed. The relationship between PCM and PHM is illustrated in Figure 3-6. PCM can be used to estimate hazard functions of mechanical components or systems in cases of sparse or no historical failure data, provided that the covariates are proportional to the hazard. A suggested research direction was to use PCM for reliability prediction based on both failure/covariates data and online CM data.

Wang [95] and Wang and Zhang [96] stated that models like PHM and PCM only use the current asset condition information (or functions of current information), rather than the whole monitoring history, to predict future asset
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

health. Therefore, a conditional residual time distribution was modelled to predict the residual life distribution of rolling element bearings based on the stochastic filtering theory [106]. All the test bearings were initially assumed to follow the Weibull delay time distribution model, and as more CM information became available, the distribution was updated. However, this model required the determination of a threshold level to indicate defect initiation point, which is hard to identify and seldom recorded in practice.

Cempel et al. [97] defined reliability in “symptom (S)” (condition data) domain, assuming the symptom of a set of continuously operating units was measured periodically and had a uniform limit value $S_r$ and breakdown value $S_b$

$$ R(S) = \Pr(S_b > S \mid S < S_r) $$

Following that definition, hazard function was defined as the relative number of units reaching breakdown value per unit increment of symptom. A logistic vector containing covariates such as system foundation stiffness and running load was also introduced into the reliability model to permit more precise determination of the system condition and symptom limit value. The model can only be validated when a sufficiently large database of real-life symptoms and covariates are available. Methods for assessing the covariate function needed for logistic vector implementation also remain an open question.

All of the models reviewed in this subsection combine reliability and condition information in failure prediction. They generally produce longer-range failure forecasts than models that only use individual asset condition information.
3.2 Remaining Challenges in Rotating Machinery Prognostics

Section 3.1 comprehensively summarised current models for predicting rotating machine failures. This section discusses how the previous studies, while they have aided the advancement of the discipline, have made only a limited contribution to developing an effective prognostic model. The candidate has identified four remaining challenges in rotating machinery prognostics, which will be discussed in the following subsections.

3.2.1 Use of Censored Event Data and Censored CM Data

Many prognostic models, especially data-driven models, require abundant failure histories. In practice, however, industrial and military communities would rarely allow their assets to run to failure. Fault detection and diagnostics technologies have resulted in considerable reductions in the number of failures due to possible early detection of faults by condition monitoring [107]. Most of the time, once a defect is detected in a unit, the unit is replaced or overhauled before it fails. Therefore, the cut-off point at which the unit will cease to function is not always known or recorded. It is only known that the unit has survived up to the time of replacement or repair but there is no information as to when it would have failed if left undisturbed. Data of this sort are called censored or suspended data. None of the existing prognostic models found in the literature has directly modelled suspended CM data.

When all the historical suspension times are treated as historical failure times, the prognostic model will produce biased estimates (underestimation) of the time to failure. Since, in many instances, a unit is replaced/overhauled once a fault is detected, treating the replacement/overhaul times as failure times defeats
the purpose of prognosis because it is the duration the failing unit can survive beyond this point that is of interest (Figure 3-7). For example, after a fault has been detected in a pump bearing, maintenance personnel have to determine whether or not the faulty bearing can survive until the end of the batch production or until the next maintenance opportunity. It is not uncommon that a machine component’s remaining useful life (from the point where a defect is detected) is substantially more than its nominal life. A prognostics specialist’s goal is to recommend a maintenance schedule that does not interrupt production or wastefully replace units that still have useful remaining life. It is the ability to estimate this remaining lifetime that is critical to optimal maintenance scheduling.

Figure 3-7: Timeline of the operational life of a machine component

On the other hand, prediction models that carefully omit suspended data from the training examples will worsen the problem of data unavailability. Degradation data are already scarce due to irregular data recording and/or the huge amount of time they take to accumulate. For example, a bearing may last several years – even under harsh operating conditions. Therefore, a good prognostic model must be able to maximise the use of available data.
3.2.2 Integration of Reliability and Prognostics

As mentioned in Section 2 of this review, reliability models based solely on event data have been reasonably well developed for machinery life estimation. However, these approaches only provide general or average estimates for the entire population of identical units to facilitate time-based maintenance. Maintenance or repair at pre-established intervals tends to incur even higher scheduled downtime. Besides, failure behaviour of each unit is a function of changes in work schedule, operating conditions and interaction between components. Therefore, current condition of an operating unit needs to be monitored online. Nevertheless, CM data, while reflecting the state of individual operating units, do not replace reliability data that reflect population characteristics. CM data mainly provide information for short-term condition prediction only. Several data-driven prognostic models [76-81] have enabled machine prognosis using time series prediction. These models mainly perform single-step-ahead predictions to estimate the vibration signal feature value at the next immediate time step. These techniques require further research because prognosis with such a short prediction horizon is not of much help for optimal maintenance scheduling. Sufficiently long lead time is often needed for effective and economical preparation of spares and human resources. A longer prediction horizon is also necessary for deciding whether a unit will last until the next maintenance opportunity (such as the end of a batch production or the next scheduled inspection).

Therefore, other than the unit-specific, online condition information measured, the prediction of remaining useful life should also be dependent on the operational age or time that the unit has survived, as well as on the general characteristics of the population to which the unit belongs. For example, given the same condition information, a bearing that is in its early operating period and is from a family with a long nominal life would be expected to have a longer remaining useful life than a bearing that is in the wear-out period from a family with a much shorter nominal life. Although several models have
attempted to use CM information for reliability estimation (as discussed in Section 3.1.2), the integration of CM information with reliability analysis has not been well explored. Some of these models are theoretical formulations that have not been validated with real life data, while others rely on an assumed failure distribution, such as Weibull distribution.

3.2.3 Non-linear Relationship between CM Indices and Actual Health of Monitoring Unit

CM data are commonly taken to indicate the health of a monitored unit. However, the measured condition indices do not always deterministically represent the actual health of the monitored unit. The challenge here lies in developing prognostic models that recognise the non-linear relationship between a unit’s actual survival condition and the measured CM indices.

Most of the existing prognostic techniques use CM indices to represent the health of the monitored unit and then use methods such as regression or time series prediction to estimate the unit’s future health, or rather, the future CM indices. In these techniques, a threshold for the CM data is predefined to represent a failure (as depicted in Figure 3-8). The prediction accuracy then relies strongly on the assumption that failure takes place at the instant of time when the relevant CM index exceeds the predetermined threshold. However, it is not uncommon that a system fails when its condition measurement is still below a predefined failure threshold or is (even) temporarily decreasing. Conversely, a system may still be performing its required function when its condition measurements already fall outside the tolerance range. Missed alarms and false alarms are common issues in practical applications of prognostic systems. Several methods have been proposed for determining thresholds for fault detection based on mathematical models, instead of solely on the asset suppliers’ suggestions or maintenance personnel’s past experiences. Decoste [108] proposed a dynamic threshold approach to set limits that can adapt to
different normal baseline conditions. The variable threshold was modelled by parametric functions which can be learnt and adapted from historical condition trending data. In the field of prognostics, effort is needed to develop a prognostic model that can deduce the non-linear relationship between a unit’s actual survival condition and the measured CM indices.

![Figure 3-8: Assumption that failure takes place at the instant of time when the condition parameter exceeds a predefined threshold](image)

3.2.4 Accuracy of Assumptions and Practicability of Requirements

Since prognosis involves projecting into the future – a future that cannot be determined with absolute certainty – assumptions and simplifications are often inevitable in prognostic modelling. Nevertheless, care must be taken to minimise these assumptions and simplifications.

Physics-based approaches can be very accurate when a correct and accurate model is available. However, the characteristics and relationships of all related components in a system and its environment are often too complicated to be modelled [100]. It has been reported that the wear of rotating machinery
components is still not fully understood at present [109, 110]. One of the main issues with empirical or physics-based models is, therefore, their inherent uncertainty due to the large number of accompanying assumptions. It was shown in numerical simulation tests and bearing life tests that physics-based models without parameter fine-tuning are unable to capture and adapt to the stochastic characteristics of defect growth [63, 64]. It was also shown that a small amount of parameter difference can result in large prediction errors with the increase of bearing running cycles. Currently, most physics-based prognostic methods focus on the prediction of crack propagation. However, in many cases, other failure modes tend to dominate and the maintenance engineer needs to correctly identify the fault type in question. Even if that has been accomplished, defect growth is not a deterministic process. Virkl er et al. [111] showed that even under well-controlled experimental conditions, crack growths of a set of identical components were vastly different. Crack growth models are also difficult to apply in practice because they require knowledge of the exact geometry and/or orientation of the crack. However, these cracks are usually very irregular and cannot be identified without disassembling the machine component. Some models have attempted to account for these uncertainties by adopting a stochastic or adaptive component such as a recursive least-squares scheme. Such adaptation techniques are similar to the adaptation in data-driven time series prediction.

Gebraeel et al. [82] based their rolling element bearing prognosis on the assumption that all bearing degradation signals possess an inherent exponential growth. It is unclear whether these complex ANN models offer an analysis more accurate than that provided by exponential curve fitting and extrapolation methods. Exponential extrapolations often have a large region of uncertainty. Due to the probabilistic nature of bearing integrity and operating condition, defect propagation rates tend to be stochastic. If the growth of a bearing defect differs from the assumed exponential pattern, the bearing degradation will be poorly extrapolated. Even for proper assumptions about the exponential defect growth, the prediction confidence interval of extrapolations often diverges to unrealistic values. Extrapolating beyond that range can lead to misleading
results. Besides, successive prediction outputs may vary vastly and seem confusing. Without probability indication, it is rather difficult for maintenance personnel to make maintenance decisions based on these prediction outputs. It is however noted that, although mathematical prognostic models have yet to prove entirely satisfactory in practice, those used for processing and managing reliability information are critical for reliability prediction work.

Consideration should also be given to finding the optimum trade-off between accuracy and costs. For example, FEA methods might be able to calculate the stress on the material surrounding a fault based on component geometry and operating conditions, but are very expensive in terms of modelling and computational time.

It is important to keep assumptions and simplifications realistic because real-life machines are complex and fault propagation is probabilistic in nature. Affordability in terms of computation time, accuracy, memory and data storage should also be considered at all stages, including model design stage.

3.3 Summary

The ability to forecast machinery failure is vital to reducing maintenance costs, operation downtime and safety hazards. This chapter synthesised the progress in the research and development of rotating machinery prognostics. The review indicates that most of the existing prognostics studies were conducted in research laboratories. It is often easy to neglect certain practical considerations when developing models in laboratory environments, where access to insights into real-life situations is often lacking. Nevertheless, the ultimate goal remains: to establish reliable prognostic systems that can be applied in real-life situations and benefit industry. To realise the greatest economic and social benefits, it is important to design every aspect of a prognostic system by considering the asset manager’s perspective.
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

The literature review indicates that there is not yet a machinery prognostic model that:

1. directly models and fully utilises suspended CM histories
2. effectively integrates population characteristics into prognostics for longer-range prediction in a probabilistic sense
3. deduces the non-linear relationship between measured condition data and actual asset health, and
4. involves minimal assumptions and requirements.

(This review has been published in the journal of Mechanical Systems and Signal Processing [112].)

The thesis proceeds to discuss a novel approach to resolve the identified limitation in the field of machinery prognostics. This approach has been published or accepted for publication in journals and conference proceedings [113-115].
CHAPTER 4  DESIGN OF A NOVEL PROGNOSTIC MODEL

The literature review in Chapter 3 revealed that the existing prognostics literature has not presented any non-parametric model that fully utilises suspended CM histories, integrates population reliability information into condition-based prediction, and derives the relationship between actual asset health and condition measurements. This chapter presents a novel approach for addressing the identified unresolved issues. The proposed model takes advantage of statistical models’ ability to extract population characteristics and reliability, and of ANNs’ ability to model the non-linear relationships between an asset’s future health and measured condition data. While the proposed model also employs an ANN as one of its tools, it differs from the other ANN-based prognostic models in several respects. Its prediction outputs are interpreted as probabilities, which collectively form a forecasted survival curve for the monitored unit. Suspended datasets are fully utilised and amalgamated directly into the training set, not by using an artificial cut-off time, but rather by using the probability that the corresponding units have not failed before a certain time as the training target. The proposed model also maximises the use of event records that do not come with corresponding CM data. Lastly, it extracts population characteristics information from the available historical CM histories and integrates this information into prognosis.

To ensure better understanding of the proposed prognostic model, background information on techniques and theories used will first be given in Section 4.1. The
architecture and working principles of the proposed model are then detailed in Section 4.2.

4.1 Enabling Techniques for the Development of Proposed Model

The proposed model combines two statistical methods and an AI method. These three methods are discussed in the following three subsections respectively. Each subsection will introduce the basic concept in the respective field, summarise the main methods and justify why a particular method is selected for use in the prognostic model.

4.1.1 Artificial Neural Networks

Artificial intelligence (AI) techniques have been widely used in numerous disciplines, including CBM. AI is the computerised approach that employs knowledge, reasoning and/or self-learning to enable machines to perform tasks which humans perform using their intelligence. It can be used for taking over some of the menial and tedious tasks in CBM, formerly performed by human experts that are often scarce. Its potential value can be better understood by contrasting it with natural, or human, intelligence as follows [116]:

- AI is more permanent. Natural intelligence is perishable from a commercial standpoint in that workers can change their place of employment or forget information. AI, on the other hand, remains as long as the computer programs have not been removed.

- AI is easier to duplicate and disseminate. Transferring a body of knowledge from one person to another is typically a lengthy process of
apprenticeship. Even so, expertise is rarely duplicated completely. Knowledge embodied in a computer program, however, can be easily transferred.

- AI can be less expensive in the long run. In many cases, computer services cost less than human power in performing the same tasks.

- AI is consistent and thorough. Natural intelligence can be erratic because humans do not always perform consistently.

- AI can be documented. Decisions made by a computer program can easily be documented by tracing the activities of the system or software. Natural intelligence is difficult to reproduce. For example, a person may reach a conclusion but, at some later date, fail to re-create the reasoning process that led to that conclusion or to even recall the assumptions that accompanied the decision.

Natural intelligence does have several advantages over AI:

- Natural intelligence is creative, whereas AI is rather uninspired. The ability to acquire knowledge is inherent in human beings; however, with AI, tailored knowledge must be built into a carefully constructed system.

- Human reasoning is able to make use of a wide context of experience at all times and bring that to bear on individual problems. AI systems, on the other hand, typically gain their power by having a very narrow focus.

AI can be divided into two categories: Symbolic Intelligence and Computational Intelligence. Symbolic Intelligence consists of knowledge-based systems or expert systems. These systems are computer programs that make decisions or solve problems in a particular field by using knowledge and analytical rules defined by human experts in the field. Symbolic Intelligence is sometimes called Good Old-Fashioned Artificial Intelligence (GOFAI) because it is based on the
assumption that thinking is nothing but symbol manipulation [117]. As an alternative to GOFAI, Computational Intelligence was proposed. It is a rather physiologically oriented approach, which relies on heuristic algorithms such as in ANNs, fuzzy systems and evolutionary computation.

The Computational Intelligence technique employed in the proposed prognostic model is the ANN method. ANNs are composed of simple elements operating in parallel. The network function is determined largely by the connections between elements. A network can be trained to perform a particular function by adjusting the values of the connection weights between elements. ANNs are an important technique in the field of AI. The development of ANNs is inspired by some of the characteristics of human brains. The human brain consists of a large number of highly connected elements called neurons [118]. Each neuron receives electrical signals through its nerve fibres (dendrites). The cell body sums and thresholds these incoming signals. These signals are then carried from the cell body out to other neurons through a single long fibre known as an “axon”. The point of contact between an axon of one cell and a dendrite of another cell is called a “synapse”. It is the arrangement of neurons and the strengths of the individual synapses that establish the function of the neural network. Learning is viewed as the development of new connections between neurons or the strengthening/weakening of synaptic junctions. The brain is a highly complex, non-linear and parallel information processing system [119]. Even though biological neurons are very slow when compared to electrical circuits, the brain is able to perform many tasks much faster than any conventional computer. This is in part due to the massively parallel structure of biological neural networks. ANNs are simplified abstractions of biological neural networks. Similar to the brain (though much less complex), the building blocks of an ANN are simple computational devices that are highly interconnected, and the connections between neurons determine the network function. The ability of ANNs to approximate arbitrary function by “learning” from observed data can be useful in prognosis, especially in cases where historical data are easier to obtain than comprehensive system physics. An eligible ANN may be able to learn the non-linear relationship between the measured series of CM information and the
health of the monitored unit, and ultimately forecast the unit’s future survival conditions or remaining useful life.

This subsection is further divided into four parts: (i) ANN training, (ii) ANN learning paradigms, (iii) types of ANNs and (iv) reasons for FFNN’s selection for use in the proposed prognostic model.

(i) ANN Training

The characteristic of ANNs that has attracted the most interest is their learning capability. ANNs can be trained to perform certain tasks. The training of ANNs consists of two phases: (a) processing phase and (b) learning phase.

(a) Processing phase

At the processing phase, an input vector is fed through the input nodes to the hidden nodes. Each hidden node performs a weighted sum of its inputs, applies an activation function to this aggregation, and passes the result to the output layer. Assuming there are \( l \) layers in an ANN, with \( l \) and \( l-1 \) denoting the upper and lower layers respectively, the result or output of a node in layer \( l \) is obtained in the following fashion [76]

\[
x_i^{(l)} = f \left( \sum_{j=1}^{n} w_{ij} x_j^{(l-1)} + b_j^{(l-1)} \right)
\]

where \( j \) is the index and \( n \) is the number of nodes in layer \( l-1 \), with \( l \leq j \leq n \), \( i \) is the index of nodes in layer \( l \), \( w_{ij} \) is the weight connecting a node \( i \) in layer \( l \) to a node \( j \) in layer \( l-1 \), \( x_j^{(l-1)} \) is the activation of node \( j \) in layer \( l-1 \), \( b_j^{(l-1)} \) is the bias and \( f \) is the transfer function. A transfer function is a mathematical representation of the relation between a neuron’s net output and its actual output. There are many different transfer functions, with the three most commonly used ones being the hard-limit transfer function, linear transfer function and log-sigmoid transfer function [120].
(b) Learning phase

After the processing phase, learning takes place. During this phase, the result or output is compared with the target, and the difference or error is propagated back through the preceding layers and the weights are adjusted based on this error.

By continuously feeding the training vectors and adjusting the weights, the error can be minimised and a relationship between the inputs and targets can be found. After the error converges, whenever relevant data are inputted to the network, the network can produce an output based on what it has learned.

(ii) ANN Learning Paradigms

There are two major learning paradigms: (a) supervised learning and (b) unsupervised learning.

(a) Supervised learning

In supervised learning, the training data consist of various pairs of input/output training examples. The aim is to find a function that matches the examples to infer the mapping implied by the data. A cost function is used to minimise the average error between the network's output and the target value over all the sample pairs. Tasks that employ supervised learning include pattern recognition (also known as classification) and regression (also known as function approximation). The supervised learning paradigm is also applicable to sequential data (e.g., for speech and gesture recognition).

(b) Unsupervised learning

In unsupervised learning, the network adapts purely in response to its inputs. Such networks can learn to pick out structure in their input. The cost function to be minimised can be any function of the input and output, depending on the task (what is to be modelled) and the a priori assumptions (the implicit properties of the model, its parameters and the observed variables). Tasks that employ
unsupervised learning are, in general, estimation problems. The applications include clustering, statistical distribution estimation and filtering.

(iii) Types of ANNs

When used without qualification, the term “Artificial Neural Network” (“ANN”) or simply “Neural Network” (“NN”) usually refers to a feed-forward neural network (FFNN). However, ANNs do not constitute one network, but a diverse family of networks. There are many other types of neural networks, including recurrent neural networks (RNNs), self-organising maps (SOMs), radial basis function networks (RBFNs), learning vector quantisation (LVQ), probabilistic neural networks (PNNs), general regression neural networks (GRNNs), cascade correlation networks, functional link networks, Gram-Charlier networks, Hebb networks, heteroassociative networks and hybrid networks. The overall function or functionality achieved is determined by the network topology, the individual neuron characteristics, the learning or training strategy and training data. This section will only discuss some of the most commonly used types of ANNs, namely: (a) FFNNs, (b) RNNs, (c) RBFNs, and (d) SOMs.

(a) Feed-Forward Neural Networks

FFNN is the first and simplest type of ANN and has arguably the widest application. In an FFNN, the information only moves in the forward direction: from the input nodes (neurons), through the hidden nodes to the output nodes. Figure 4-1 illustrates a fully connected FFNN with one hidden layer. The universal approximation theorem for neural networks states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by an FFNN with just one hidden layer [121]. This result holds only for restricted classes of transfer functions, e.g., for the sigmoidal functions.
(b) Recurrent Neural Networks

Contrary to an FFNN, an RNN is a network with bi-directional data flow; some of the hidden unit activations or output values are fed back into the network as inputs. Figure 4-2 illustrates an RNN with three input nodes, five hidden nodes, one output node and, in addition, a feedback loop linked from the output node to an extra input node. RNNs are potentially more powerful than FFNNs, due to their ability to recognise and recall temporal and spatial patterns. However, the behaviour of RNNs is much more complex than that of FFNNs [118]. For a given input and a given initial network output, the response of the network may converge to a stable output. However, it may also oscillate, explode to infinity, or behave chaotically.
(c) Radial Basis Function Networks
RBFN is a feed forward network whose hidden layer uses non-linear radial basis functions as activation functions (Figure 4-3). Various functions have been tested as activation functions for RBFNs. In time series modelling, the most common activation function is the thin-plate spline [122]. In pattern classification, the Gaussian function is preferred [123-126]. RBF hidden layer units have a receptive field which has a centre: a particular input value at which the units have a maximal output. Their output tails off as the input moves away from this point. The centres and the standard deviations of the activation functions need to be decided by examining the vectors in the training data.

Figure 4-2: Recurrent neural network
(d) Self-Organising Maps
SOM is a type of ANN that is trained (using unsupervised learning) to produce a low-dimensional (typically two-dimensional), discretised representation of the input space of the training samples called a “map” (Figure 4-4). It seeks to retain the topological ordering of the input and is useful for visualising low-dimensional views of high-dimensional data. The training uses competitive learning. When a training input is fed to the SOM, its Euclidean distance to all weight vectors is computed. The neuron with weight vector most similar to the input is called the best matching unit (BMU). The weights of the BMU and neurons close to it are adjusted towards the input vector. The magnitude of the change decreases with time and with distance from the BMU. This process is repeated for each input vector for a (usually large) number of cycles. The network winds up associating output nodes with groups or patterns in the input data set.

Figure 4-3: Radial basis function network
(iv) Reasons for Choosing Feed-Forward Neural Network

Numerous studies across various disciplines have demonstrated the merits of ANNs, including their ability to (a) perform faster than system identification techniques in multivariate prognosis [127]; (b) perform at least as well as the best traditional statistical methods without requiring untenable distributional assumptions [76, 128]; and (c) capture complex phenomena without a priori knowledge [129].

Although ANNs have been applied successfully to a variety of real-world scenarios, they are built for accuracy in complex modelling and not for reasoning transparency. As a result, ANNs have been criticised for lacking documentation of decision-making in a trained network. However, it was argued in Ref. [129] that the view of ANNs being less transparent than traditional statistical models misunderstands the situation. ANNs are only less transparent when the phenomenon is complex and the networks have to be sufficiently complex to represent it. However, if the phenomenon is linear, for example, then a two layer (no hidden layer) ANN with linear transfer functions is mathematically identical to linear regression, and the network weights are identical to the beta coefficients in linear regression. Increase in model
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

complexity reduces the transparency of both traditional statistical models and ANN models. The difference is simply that ANNs are more capable of modelling complex phenomena and, consequently, need a more complex structure to represent these phenomena. Various mechanisms that automatically compile ANNs into symbolic rules for overcoming this practical limitation have been proposed, such as those reviewed in Ref. [130]. Rule extraction will not be discussed in this paper since the objective here is not to comprehensively explain how the trend in a given series of condition indices affects an asset’s reliability.

As Stergiou and Siganos [131] pointed out, ANNs do not perform miracles, but they can produce some amazing results if used sensibly. ANNs, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections when given new situations of interest, and to answer "what if" questions.

Other advantages include:

- Adaptive learning: An ANN has the ability to learn how to do tasks based on the data provided for training or initial experience.
- Self-organisation: An ANN can create its own organisation or representation of the information it receives during learning.
- Real time operation: ANN computations can be very fast as they are carried out in parallel, and special hardware devices which take advantage of this capability are being designed and manufactured.
- Fault tolerance via redundant information coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage.
Non-linear information processing: ANNs are non-linear information processors and hence have high capabilities of approximation, classification and noise-immunity.

An FFNN with sigmoidal transfer functions is selected for use in the proposed prognostic model. As mentioned, the universal approximation theorem for neural networks states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by an FFNN with just one hidden layer [123]. This result holds only for restricted classes of transfer functions, and sigmoidal functions are one of these classes.

RNNs are potentially more powerful than FFNNs, due to their ability to recognise and recall temporal and spatial patterns. However, the behaviour of RNNs is much more complex than that of FFNNs [118]. For a given input and a given initial network output, the response of an RNN may converge to a stable output. However, it may also oscillate, explode to infinity, or behave chaotically.

RBFNs, due to their non-linear approximation properties, are able to model complex mappings, which normal FFNNs can only model by means of multiple intermediary layers. However, there is considerable difficulty in deciding the centres and the standard deviations of the RBF activation functions. The vectors in training data need to be manually examined.

SOMs are unsupervised learning networks and are more suitable for visualising high-dimensional data. FFNN, on the other hand, is a supervised learning network. The capability of supervised learning networks is significantly greater than that of unsupervised learning networks [132]. For failure prognosis, the most important factor is accuracy and the network must be trained with targets that are as close to the actual values as possible.
4.1.2 Statistical Analysis of Survival/Failure Time

A survival analysis technique called the Kaplan-Meier (KM) estimator will also be incorporated into the proposed prognostic model. This section introduces the basic concept of survival analysis, discusses several commonly-used techniques and, finally, justifies why the KM estimator is selected for the proposed model.

Survival/duration/reliability analysis involves the study of the time between some defined time origin and a predefined event. In medical research, survival analysis might be used to measure the proportion of surviving patients past a certain time after surgery. An economist might use it to measure the duration of unemployment after a job loss. An engineer might use it to measure the mean time to failure of machine components. Several reasons why survival analysis differs from standard statistical analyses include the lack of symmetry, restriction to positive values (so a normal distribution cannot be assumed), and the common involvement of incomplete data.

The terms “survival probability” and “reliability” are used interchangeably in this report. Survival probability (or reliability) is defined here as the probability that a unit (component, subsystem or system) will be capable of performing a required function for a stated period of time. If $T$ denotes the random variable representing the unit’s operating time to failure, the reliability function $R(t)$ is defined as

$$R(t) = Pr[T > t]$$

for unconditional reliability, assuming that the unit has not yet been put into operation, and as

$$R(t) = Pr[t < T \leq t + \Delta t \mid T > t] = \frac{Pr[t < T \leq t + \Delta t]}{Pr[T > t]}$$

for conditional reliability, assuming that the unit in operation has not failed until some time $t$. 

- 65 -
The rest of this subsection presents (i) classifica
tion of survival data, (ii) several
commonly used parametric failure distributions, (iii) the proportional hazards
models and (iv) the Kaplan-Meier estimator, and (v) reasons why the Kaplan-
Meier technique was chosen for the survival analysis in this work.

(i) Classification of Survival Data

(a) Complete Data
Complete data are samples with an observed or known event of interest (e.g.
time to death or failure). In the case of life data analysis, a complete dataset
would be composed of the times to failure of all units in the sample [134]. For
example, if five units were tested and they all failed (and their times to failure
were recorded), there would be a complete dataset or their condition data would
be complete condition histories. Figure 4-5 illustrates complete data.

![Figure 4-5: Complete data](image)

(b) Censored Data
A fundamental feature of survival analysis is that the survival times are
frequently censored. Reasons for censoring include: samples being withdrawn
from a study early, still surviving at the end of the study, or reaching the end event of interest due to other causes. There are three types of censoring schemes: right-censoring (also called “suspension”), interval-censoring and left-censoring. **Suspension** is the most common case of censoring. In the case of asset life data, suspended datasets are composed of units that did not fail. Continuing with the previous illustration, if only three of the five tested units had failed by the end of the test, there would be suspended data for the two unfailed units (see Figure 4-6). The term “right-censored” implies that the event of interest (e.g. death or failure) is to the right of our data point.

![Figure 4-6: Right-Censoring / Suspension](image)

The second type of censoring, **interval-censoring**, reflects uncertainty as to the exact failure time within an interval. For example, if five units were left running and were only inspected periodically, there would only be information as to whether a unit failed or did not fail in the interval between inspections. Figure 4-7 illustrates interval-censoring.
In left-censored data, it is only known that failure occurred before a certain time. Using a similar illustration as an example, if it was known that Unit 1 failed sometime before 600 hours and Unit 3 failed sometime before 500 hours, but exactly when is not known, the two samples would be left-censored (see Figure 4-8). This is similar to interval-censoring with interval starting at $t = 0$.

![Figure 4-7: Interval-Censoring](image)

![Figure 4-8: Left-Censoring](image)
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

In the case of condition-based maintenance, failure is usually reported immediately and rarely detected only at the next inspection. Censoring also generally happens some time after (to the right of) the last data point. Therefore, only right-censoring/suspension is considered in this work.

(ii) Parametric Failure Distributions

Failure distribution, $F(t)$, is a cumulative distribution function that describes the probability of failure prior to time $t$,

$$F(t) = 1 - R(t) \quad (4-4)$$

The failure distribution is the integral of the failure probability density function (PDF), $f(x)$,

$$F(t) = \int_{0}^{t} f(x) dx \quad (4-5)$$

This subsection introduces several failure distributions commonly used in reliability analyses. They include the (a) exponential distribution, (b) Weibull distribution and (c) log-normal distribution.

(a) Exponential distribution

This distribution is used to model the behaviour of units that have a constant failure rate. Mathematically, it is a fairly simple distribution, which sometimes leads to its use in inappropriate situations. The PDF of an exponential distribution has the form [105]

$$f(t) = \lambda e^{-\lambda t} \quad (4-6)$$

where $\lambda > 0$ is the failure rate parameter. The distribution is supported on the interval $[0, \infty)$. The corresponding reliability function is given by

$$R(t) = e^{-\lambda t} \quad (4-7)$$
(b) Weibull distribution

The Weibull distribution is among the most popular in the field of reliability analysis because it is able to accommodate various types of behaviour, including infant mortality in the “bath-tub” curve. A typical Weibull PDF is defined by

\[
f(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta-1} e^{-\left( \frac{t}{\eta} \right)^\beta}
\]

for \( t \geq 0 \) and \( f(t) = 0 \) for \( t < 0 \), where \( \beta > 0 \) is the shape parameter and \( \eta > 0 \) is the scale parameter of the distribution. Figure 4-9 shows some sample Weibull PDF with varying shape parameters.

![Typical Weibull Distributions](image)

Figure 4-9: Weibull probability density function with different shape parameters

The Weibull reliability function is given by

\[
R(t) = e^{-\left( \frac{t}{\eta} \right)^\beta}
\]

(4-9)

where \( \beta \) is the shape parameter and \( \eta \) is the scale parameter of the distribution.
(c) Normal distribution
The normal distribution is used when the distribution of purely random events is assumed to be normal. This assumption is related to the central limit theorem [135], which states that the mean of any set of variables with any distribution having a finite mean and variance, tends to the normal distribution. The PDF for normal distribution is given by

\[ f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \]  

(4-10)

where \( \sigma > 0 \) and \( \mu \) are the standard deviation and the mean of the distribution respectively.

(d) Log-normal distribution
The log-normal distribution is the probability distribution of any random variable whose logarithm is normally distributed. It has the PDF

\[ f(t) = \frac{1}{t\sigma \sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}} \]  

(4-11)

for \( x > 0 \), where \( \mu \) and \( \sigma \) are the mean and standard deviation of the variable’s logarithm respectively. The log-normal reliability function is

\[ R(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \]  

(4-12)

with \( \Phi(z) \) denoting the standard normal cumulative distribution function.

(iii) Proportional Hazards Models
The PHM has been widely used in both medical survival analyses and machinery reliability studies. It can be written as

\[ h_X(t) = h_0(t)e^{\beta x} \]  

(4-13)
where \( h_0(t) \) is the baseline hazard function. The transpose of vector \( X \) contains the \( p \) covariates of the respective unit, \( X^T = (x_1, x_2, \ldots, x_p) \); whereas vector \( \beta \) contains the coefficients for each of the \( p \) covariates, \( \beta^T = (\beta_1, \beta_2, \ldots, \beta_p) \).

PHMs assume that hazard changes proportionately with covariates and that the proportionality constant is the same at all times. By further assuming a particular failure distribution (such as a Weibull distribution or the exponential distribution), a fully parametric model for proportional hazards is obtained. The coefficients \( \beta_1, \beta_2, \ldots, \beta_p \) are estimated by the method of maximum likelihood [136].

There were many advocates for the distribution-free (nonparametric) modelling of data. Cox’s PHM [105] was then introduced to address this issue. It is the most general regression model because it is not based on any assumptions concerning the nature or shape of the underlying failure distribution. This method allows the baseline hazard estimation to be deferred until after the regression coefficients \( \beta_1, \beta_2, \ldots, \beta_p \) are obtained. However, Cox’s PHM still relies on the proportionality assumption, which is the assumption that the ratio of hazard functions for two observations with different covariate values is the same at any time \( t \).

**(iv) Kaplan-Meier Estimator**

The KM estimator, developed by Kaplan and Meier [137], is an empirical nonparametric method for estimating a distribution function from life-time data that may be suspended.

Let \( t_j \) denote the failure time of each of the \( m \) historical units under study, and \( j = 1, 2, \ldots, k \) (\( k = m \) if all the historical units failed at different times). Also, let \( d_j \) denote the number of deaths or failures that occur at \( t_j \), and \( n_j \) denote the corresponding number of historical units remaining at risk. Note that \( n_j = n_{j-1} - d_{j-1} \). Then, loosely speaking, the survival probability beyond time \( t_j \)
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

depends conditionally on the survival probability beyond time \( t_{j-1} \) [138]. A rule that can be used to determine conditional probability is

\[
P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (4-14)
\]

where

\( P(A | B) \) is the conditional probability that event A will occur given that event B has already occurred;
\( P(A \cap B) \) is the unconditional probability that event A and event B both occur;
\( P(B) \) is the unconditional probability that event B occurs.

From equation 4-14, the probability that event A and event B both occur is

\[
P(A \cap B) = P(B) \times P(A | B) \quad (4-15)
\]

Let \( A \) be the event of surviving the first time interval and \( B \) be the event of surviving the following interval. \( A \cap B \) is then the event of surviving both intervals. For any time \( t \in [t_1, t_2) \), the survival probability is

\[
S(t) = \Pr(surviving[0,t_1]) \times \Pr(surviving[t_1,t] | surviving[0,t_1])
\]

\[
\hat{S}(t) = 1 \times \left( n_1 - d_1 \right) n_1 \quad (4-16)
\]

or

\[
\hat{S}(t) = 1 \times \left( 1 - \frac{d_1}{n_1} \right) \quad (4-17)
\]

Similarly, for any time \( t \in [t_2, t_3) \), the survival probability is

\[
S(t) = \Pr(surviving[t_1,t_3]) \times \Pr(surviving[t_2,t] | surviving[t_1,t_3]) \quad (4-18)
\]
By using this recursive idea, the estimate of the survival function can be written as

\[
\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right) \times \left(1 - \frac{d_2}{n_2}\right) \ldots \left(1 - \frac{d_k}{n_k}\right) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j}\right)
\]  

(4-20)

This method is known as the KM estimation of survival or reliability function.

As mentioned at the beginning of this subsection, the KM estimator can account for suspended data. In situations where some of the failure times are not known (e.g. preventive replacement of machine components before failure), the KM estimator can be used to produce the most accurate possible survivor/reliability function, taking into account all information available. The purpose here is to allow a unit to contribute to the survivor function for the entire length of time it was monitored, and to then statistically "remove" the unit from the function after that. For example, if a rolling element bearing had been monitored for two years and was still operating when it was preventively replaced during an overhaul, the fact that this bearing had survived for two years should contribute to the reliability information. We should not count the bearing as having failed after two years, and we should not count it as having survived longer than that either. In the procedure of KM estimation, after a unit is suspended, it is no longer considered as one of the units at risk and, hence, no longer contributes to the number of remaining units (\(n_j\) in Equation 4-20).

A plot of the KM survivor function is a series of horizontal steps of declining magnitude which, when a large enough sample is taken, approaches the true survival function for that population. A sample KM plot is shown in Figure 4-10. Each tick mark represents a suspension. Each time one or more units fail, the curve takes a step down. When a unit is suspended, the curve does not take a step down. However, because the suspension reduces the number of units...
contributing to the curve, each failure after that point represents a higher proportion of the remaining population, and so every step down afterwards will be a little bit larger than it would have been.

Figure 4-10: A sample Kaplan-Meier plot

(v) Reasons for Choice of the Kaplan-Meier Estimator

The paper by Kaplan and Meier [137] that proposed the KM estimator is one of the top five most referenced papers in the field of sciences. The KM estimator has been a widely used survival analysis technique in medical research. It is a useful technique when the number of cases is small but representative, and the exact survival times are known. It has an advantage over competing methods in that it does not depend on a choice of intervals. Methods such as the life table analysis work with fixed length time intervals. Information is lost when failure or suspension events are grouped in fixed intervals of time. The wider the intervals, the more information is lost.
Another source of difficulty in the analysis of lifetime data is the possibility that some units may not be observed for the full time to failure. For example, not all components may have failed at the close of a life-testing experiment in industrial reliability. Periodic overhauls of machinery also cause some components to be preventively replaced before they fail. As mentioned earlier in this section, the KM estimator can account for suspended data and allow the suspended samples to contribute to the reliability function and population characteristics information.

The various failure distributions (exponential, Weibull, normal and log-normal distributions) as well as the PHM, are parametric techniques; they require specification of the functional form of the distribution that failure time would have in the absence of suspension. The Cox’s PHM is a technique that does not require the assumption of an underlying distribution. However, Cox’s PHM does use the proportionality assumption, which is not appropriate in many applications. Inferential consequences of using the model when the assumption does not hold can be high. Carter et al. [139] demonstrated the high consequences by fitting a PHM to a set of gastric cancer data. Other more flexible variations able to cope with non-proportional hazards usually require non-standard estimation techniques, which are often complex and thus not favoured in application. The PHM also assumes that there is a log-linear relationship between the independent variables and the underlying hazard function. Therefore, the Cox PHM is also called a semi-parametric PHM. The KM estimator, on the other hand, is a nonparametric technique which requires no such assumptions. The ability to analyse data without assuming an underlying life distribution avoids the potentially large errors brought about by making incorrect assumptions about the distribution.

As one of the main objectives of the research is to develop a prognostic model that can account for censored data and requires minimal assumptions, the KM estimator is selected for use in the proposed model.
4.1.3 Kernel Probability Density Estimation

Another enabling technique used in the proposed model is the kernel probability density estimation (PDE). This subsection will first present the basic concept of PDE. Several PDE models, including the kernel estimator, will be introduced in Subsection (i). Finally, the selection of the kernel model as an ingredient of the proposed prognostic model will be discussed in Subsection (ii).

PDE is the construction of an estimate of the PDF from observed data. Consider any random quantity \( X \). Knowing its PDF, \( f \), allows the calculation of probabilities associated with \( X \) \[ \int_a^b f(x)dx \text{ for all } a < b \] (4-21)

One approach to density estimation is parametric. Assume that data are drawn from a known parametric such as the normal or Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \). The underlying probability density can then be estimated by finding the estimates of \( \mu \) and \( \sigma^2 \) from the data and substituting these estimates into the Gaussian density formula. This work will not be considering parametric estimators: the focus will be on nonparametric methods, which require less rigid assumptions about the distribution of observed data.

(i) Main existing estimators

Several important methods for estimating probability density will be discussed here. They include (a) histograms, (b) the kernel estimator, (c) the nearest neighbour estimator, (d) the variable kernel method and (e) the maximum penalised likelihood method. It is convenient to define some standard notation. It will be assumed that we are given a sample of \( n \) real observations \( X_1, \ldots, X_n \) whose underlying density is to be estimated. The symbol \( \hat{f} \) denotes the corresponding density estimator being considered.
(a) Histograms

The oldest and most widely used density estimator is the histogram. Figure 4-11 shows a sample histogram.

Given an origin $x_0$ and a bin width $h$, the bins of the histogram are defined to be the intervals $[x_0 + nh, x_0 + (n+1)h)$ for integers $n$. The histogram is then defined by

$$
\hat{f}(x) = \frac{1}{nh} \left( \text{number of } X_i \text{ in same bin as } x \right)
$$

(4-22)

Note that an origin and a bin width have to be selected before the histogram can be constructed. The amount of smoothing inherent in the procedure depends primarily on the choice of bin width. The histogram can be generalised by allowing the bin widths to vary.

(b) The kernel estimator

The kernel density estimator can be defined as

$$
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)
$$

(4-23)
where \( h \) is the bin width and \( K \) is some kernel function and is usually taken to be a standard Gaussian function that satisfies the condition

\[
\int_{-\infty}^{\infty} K(x)dx = 1 \quad (4-24)
\]

The kernel estimator can be considered as a sum of “bumps” placed at each observation. An illustration is depicted in Figure 4-12 [142], where six Gaussian “bumps” (broken lines) are shown as well as their sum (solid line). The kernel function determines the shape of the bumps while the window or bin width \( h \) determines their width. Apart from the histogram, the kernel estimator is probably the most commonly used estimator. However, when it is applied to data from distributions with long tails, spurious noise may appear in the tails of the estimates because its window width is fixed across the entire sample. If the estimates are smoothed sufficiently to avoid the under-smoothed tails, essential detail in the main part of the distribution may be masked.

(c) The nearest neighbour estimator
The nearest neighbour class of estimators attempt to adapt the degree of smoothing to the local density of data. The degree of smoothing is controlled by an integer \( k \). If the distance \( d(x, y) \) between two points on the line is defined to be \( |x - y| \), and \( d_1(t) \leq d_2(t) \leq \ldots \leq d_n(t) \) is defined to be the distances, arranged in ascending order, from \( t \) to the points of the sample, the \( k \)th nearest neighbour density estimate is then written as [140]
As the distance $d_k(t)$ in the tails of the distribution will be larger than in the main part of the distribution, the problem of under-smoothing in the tails should be reduced.

(d) The variable kernel estimator
The variable kernel method is also aimed at adapting the degree of smoothing to the local density of data. The estimate is constructed similarly to the classical kernel estimate, but the scale parameter of the “bumps” placed on the data points is allowed to vary from one data point to another. If $d_{j,k}$ is the distance from $X_i$ to the $k$th nearest point in the set comprising the other $n-1$ data points, the variable kernel estimate is defined by

$$
\hat{f}(t) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_{d_{j,k}}} K\left( \frac{t - X_j}{h_{d_{j,k}}} \right)
$$

(4-26)

since the window width of the kernel placed on the point $X_j$ is proportional to $d_{j,k}$, kernels in regions with sparse data will be flatter.

(e) Maximum penalised likelihood estimators
The likelihood of a curve $g$ as density underlying a set of identically distributed observations is given by

$$
L(g | X_1, ..., X_n) = \prod_{i=1}^{n} g(X_i)
$$

(4-27)

This likelihood has no finite maximum over the class of all densities. Let $\alpha$ be a positive smoothing parameter and $R(g)$ a functional that quantifies the roughness of $g$, the penalised log likelihood is written as

$$
\begin{align*}
\hat{g}(x) &= \arg\max_{g \in \mathcal{G}} \{ L(g | X_1, ..., X_n) - \alpha R(g) \}
\end{align*}
$$

(4-28)
The log likelihood term \( \sum \log g(X_i) \) measures how well \( g \) fits the data. The penalised log likelihood can thus be considered as a way of quantifying the conflict between smoothness and goodness-of-fit to the data. The PDE, \( \hat{f} \), is said to be a maximum penalised likelihood density estimate if it maximises \( l_\alpha(g) \) over the class of all curves \( g \) that satisfy \( \int_{-\infty}^{\infty} g = 1 \), \( g(x) \geq 0 \) for all \( x \), and \( R(g) < \infty \).

(ii) Reasons for choice of kernel probability density estimator

Parametric approaches, such as the Gaussian method and the Weibull method, require rigid assumptions about the distribution of the observed data. With nonparametric methods such as the kernel estimator, the data are allowed to speak for themselves in determining the estimate of probability density.

Those who are sceptical about density estimation often ask why it is ever necessary to use methods more sophisticated than the simple histogram [140]. The drawback of histograms is that they depend substantially on the context of the problem. In many cases (including the prognostic model that will be proposed later in the chapter), use of histograms leads to inefficient use of data. The discontinuity of histograms also causes extreme difficulty if derivatives of the estimates are required. Since, in this work, density estimates are used as an intermediate component of other methods, and derivatives of these estimates are required, the case for using alternatives to histograms is strong.

The nearest neighbour estimation is identical to the kernel estimation when estimating a density at a single point. It is only when estimating the density at several points, or when constructing an estimate of the entire density function, that the two methods will give different results. Unfortunately, as illustrated in Ref. [140], the overall estimates obtained by the nearest neighbour method are
not very satisfactory. They are prone to local noise and also have very heavy tails and infinite integral.

Variable kernel approach is worth considering if under-smoothing in the tails of distributions is likely to cause difficulties, which is not the case in this application. The extension of the maximum penalised likelihood approach to multivariate density estimation is possible in principle, but has not received much attention, not least because of the severe computational difficulties involved.

The kernel probability density estimator is a good choice for many practical purposes; it is simple and intuitively appealing, and its mathematical properties are well understood. Given the various merits of the kernel estimator in comparison to the other methods, the kernel estimator is selected for use in the proposed prognostic model.

4.2 Architecture of the Proposed Prognostic Model

Now that Section 4.1 has introduced the enabling techniques that will be employed in the proposed prognostic model, this section now describes how these techniques are integrated into the model architecture.

As mentioned in the previous section, the proposed prognostic model employs an FFNN as one of its tools. The FFNN used consists of one hidden layer (Figure 4-13).
The number of network input nodes will be $d + 1$, where $d$ is the number of network delays that is found to produce the best prediction output. Therefore, the input values consist of the latest condition index of the monitored unit and $d$ previous indices. Note that if inspection intervals are not equally spaced, interpolation can be carried out to evenly space data points with an interval of $\Delta$. The condition indicator should be representative of the unit health. This indicator can be a single feature (e.g., vibration RMS level) or a fusion of several features (e.g., a fusion of vibration RMS level, temperature reading and oil particle count).

The number of output nodes depends on the desired prediction horizon, $h\Delta$. Let $I$ denote probability of survival or reliable operation, $\Delta$ denote the fixed time interval between condition measurements, $n$ denote the number of future intervals that the $n$th output node represents and $n = 1, 2, 3, ..., h$. The activation of the $n$th output node is trained with, and interpreted as, $\hat{S}(t + n\Delta)$, which is the probability that the unit would survive up to the $n$th next time interval. For example, the 1st output $\hat{S}(t + \Delta)$ represents the probability of the unit surviving the 1st next interval, the
2nd output $\hat{S}(t+2\Delta)$ represents the probability of the unit surviving the 2nd next time interval and so on. Collectively, the survival probabilities represent the forecasted survival curve for the monitored unit at the time of prediction.

Conventional ANN-based prognostic models have approached machine prognosis as a time series prediction problem, where the target output is the condition value at the following measurement. Then, the time to failure is usually determined by the moment the forecasted condition values exceed an assumed failure threshold. The target outputs for the ANN used in the proposed model, however, are survival probabilities for the individual historical units at each specific measurement. These survival probabilities are computed taking into account the actual survival status of the historical unit at the time of measurement, as well as the health of this historical unit compared to the health of the entire population at a similar operating age. These two considerations that contribute to determining the training target probabilities are detailed in Sections 4.2.1 and 4.2.2 respectively.

### 4.2.1 Kaplan-Meier Estimation of Survival Probability

As mentioned, most of the existing prognostic techniques involve setting a threshold for the condition data values to represent a failure (as depicted in Figure 3-8). The prediction accuracy then relies strongly on the assumption that a failure occurs at the moment the condition indicator exceeds the predefined threshold. The proposed prognostic model, on the other hand, is trained to recognise the non-linear relationship between a unit’s actual survival condition and the acquired series of prognostic information. During training, whenever the latest and delayed condition values are fed into the ANN as input, the survival conditions in the following intervals within the prediction horizon need to be presented to the ANN as target outputs.
(i) Training Targets for Complete Histories

A trending history is considered *complete* if the monitored unit has reached failure when removed from operation. (The definition of “failure” depends on the application. As mentioned, a unit is generally considered to have failed if it has ceased to perform the required functions satisfactorily. Maintenance personnel usually decide whether a failure threshold has been reached based on past experiences.) Let $i = 1, 2, \ldots, m$, where $m$ represents the number of monitored historical units. If unit $i$ has reached failure before repair or replacement, its survival probability is assigned with a value of “1” (100% survival) up until its failure time, $T_i$, and a value of “0” thereafter.

$$ S_{KM,i}(t + k\Delta) = \begin{cases} 1, & 0 \leq t + k\Delta < T_i \\ 0, & t + k\Delta \geq T_i \end{cases} $$

Note that we consider all functions discussed here to be the true function estimated from the given data and the hat “^” is dropped for notational clarity.

(ii) Training Targets for Suspended Degradation Datasets

A trending history is considered *suspended* if the unit has not reached failure when it is overhauled or removed from operation. For such suspended histories, the survival probability is similarly assigned a value of “1” up until the time interval in which survival was last observed. Survival probability for the following time intervals is computed using a variation of the KM estimator [137] based on the true survival rate of the other complete histories from this moment onwards. The aim of using this method is to compute the most accurate survival probability possible, taking into account all of the information available. The baseline survival probability in interval $t$ is estimated by the KM estimator as follows:

$$ \hat{S}(t) = \prod_{j \in \mathcal{S}} \left(1 - \frac{d_j}{n_j}\right) $$

- 85 -
where \( d_j \) is the number of failures up to interval \( t_j \) and \( n_j \) is the number of units at risk just prior to interval \( t_j \). In this work, the KM estimator which computes overall survival probability for the total set of historical units under study was adapted to estimate the survival probability for individual units. Note that the term “overall” is used in this work to indicate that the characteristics are estimated for the whole population of historical units.

For *true suspensions* – where suspended units that are repaired/replaced due to non-deterioration factors (such as units still remaining in operation at the end of data collection, or units that are being opportunistically overhauled/replaced at operation stoppages caused by the failure of other components) – the adapted KM estimator tracks the cumulative survival probability of the suspended unit \( i \) in the following fashion:

\[
S_{KM,i}(t+k\Delta) = \begin{cases} 
1, & 0 \leq t+k\Delta < L_i \\
\prod_{L_i \leq t+j\Delta \leq t+k\Delta} \left( 1 - \frac{d_{t+j\Delta}}{n_{t+j\Delta}} \right), & t+k\Delta \geq L_i 
\end{cases}
\]

(4-31)

where \( L_i \) denotes the time interval in which historical unit \( i \) was last observed to be still surviving. Note that we use the last observed survival interval \( L_i \), as the starting time, rather than time zero, to compute the cumulative survival probabilities of each suspended unit.

For *informative suspensions* (suspended units that are repaired/replaced to prevent failures because a fault has been detected), the adapted KM estimator computes the cumulative survival probability of the suspended unit \( i \) as follows:

\[
S_{KM,i}(t+k\Delta) = \begin{cases} 
1, & 0 \leq t+k\Delta < L_i \\
\mu_i \cdot \prod_{L_i \leq t+j\Delta \leq t+k\Delta} \left( 1 - \frac{d_{t+j\Delta}}{n_{t+j\Delta}} \right), & t+k\Delta > L_i 
\end{cases}
\]

(4-32)
where $\mu_i$ is the health index estimated based on the fault severity of the unit at repair/replacement and $0 \leq \mu_i \leq 1$. For example, a unit that has a very minor fault can be given a health index of 0.7 by the maintenance personnel, whereas a unit that is considerably damaged can be given a health index of 0.2.

### 4.2.2 Failure PDF Estimation Based on Degradation Data

The previous subsection has discussed how the measured condition data can be associated with the actual unit survival status, as well as how suspended CM histories are utilised in the modelling of historical data. However, there is more information that can be extracted from the available data. The population characteristics of all available historical units, as well as the individual unit’s health in comparison to the average population health in each time interval, also contribute to the unit’s survival probability estimation.

In classical reliability theory, the population characteristics are commonly expressed through the overall probability density estimated from the all historical failure times, which, in this case, can be defined as the horizontal intersections of the condition indices with the failure threshold (see Figure 4-14). The reliability function is the probability distribution of random variable, $T$, which represents a unit’s operating time to failure, defined here as

$$S(t) = \Pr[T > t] = \int_t^\infty f(t)dt$$

where $f(t)$ is the PDF of $T$. 

\[ (4-33) \]
The classical reliability method has several deficiencies. Firstly, the classical reliability method considers only two possible states of the units: the state of functioning and the state of failure [144]. The changes in condition of units during their operating lives are not trended. Each unit is accepted as a “black box” which performs the required function until it fails. However, maintenance engineers would often like to know “what is going on inside the box” and would like to use the degradation process to forecast failures.

The second deficiency of the classical reliability approach is that it only provides an overall reliability estimate for the whole population of units. This type of estimations might be useful to manufacturers that produce units in high volumes but are of little value to end users. For example, a maintenance engineer would be more interested in the specific reliability information of a particular piece of component currently in operation, than in the mean time to failure of the whole population of such a component. Therefore, more unit-specific information is needed to schedule maintenance more optimally.
Recent advances in CM technologies have enabled the collection of data that actually reflect the condition or health state of individual units throughout their operating lifetime. This new advancement has given rise to the challenge of incorporating the unit health condition into the classical reliability calculation.

A more “condition-based” approach to estimating reliability is then based on the degradation data of historical units and their change over time [144]. All available CM histories are used to access the PDE of the condition values in each time interval (see Figure 4-15). Note that the probability density is estimated from the vertical intersections of the condition values with time $t = t_j$. It is an instantaneous population characteristic, based on the real degradation processes of the population. This is a more “engineering” approach of estimating reliability characteristics compared to the classical reliability method.
The process of change in condition values of each history can be described as a random process since it is impossible to predetermine how the condition will progress. It can be expressed as a series of regression points, each having a certain probability of occurrence.

Let $y_j(t)$ denote the random function of time which describes the random process of condition change in unit $i$, where $i = 1, 2, ..., m$ and $m$ represents the number of monitored historical units.

Let $Y(t)$ be the condition value for unit $i$ at operating age $t$, and $Y(t)$ a vector containing the condition values from all of the $m$ historical units at time $t$

$$Y(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_m(t) \end{bmatrix}$$

(4-34)
The PDE of condition values at a time step $t$ is denoted as $f(Y | t)$. The probability of the condition indices not exceeding failure threshold is defined by the following equation:

$$\Pr[Y(t) < Y_{\text{thresh}}] = \int_{0}^{Y_{\text{thresh}}} f(Y | t) dY$$

(4-35)

In the case considered, this overall probability of condition values not exceeding failure threshold is also the overall probability of survival:

$$S(t) = \Pr[Y(t) < Y_{\text{thresh}}] = \int_{0}^{Y_{\text{thresh}}} f(Y | t) dY$$

(4-36)

The above equation shows that reliability function can be estimated taking into consideration the mechanism of change in the condition of each historical unit. Such an approach provides a fuller “picture” of the degradation characteristics throughout the asset lifetime since it is based on the continuous process of condition change rather than on the overall time-to-failure approach.

The random processes discussed are continuous in theory and there are an infinite number of such distributions corresponding to the infinite number of possible instants of operating time. However, it is impossible in engineering practice to determine the condition values at every instant of time. Furthermore, condition measurements of each historical unit are not always recorded at the same operating age or time interval. One solution is to obtain the continuous function by constructing a regression line over the points of condition measurements. In this way, an infinite number of PDFs can be estimated over the operating time. However, a less computationally intensive solution is to interpolate the condition values to produce equally spaced measurement points, one in each time interval. The simplified example is given in Figure 4-16.
In this way, the probability density of condition values at each time interval can be estimated. Other reliability characteristics can also be determined. The overall probability of failure is given by

Figure 4-16: Interpolation of data points for estimating the probability density in each operating interval
Since the probability of survival and the probability of failure are mutually exclusive.

However, the reliability characteristics calculated so far are an overall estimate for the whole population of historical units. To estimate specific survival probability for each historical unit \(i\), for the purpose of generating training targets for the FFNN, we successively multiply the probability of the surviving units having condition indices higher than the observed index of unit \(i\), but lower than the threshold. (Note that we consider that the condition value, which represents degradation of the corresponding asset, will not decrease. This is an assumption that will yield us a conservative estimate of survival probability.)

Let \(k = 1, 2, \ldots\) and setting the constraint to be such that the historical condition index at \(t + k\Delta\) is known but the survival status at \(t + k\Delta\) is not, the conditional probability of a unit \(i\) surviving interval \(t + k\Delta\) is

\[
S_{PDF,i}(t + k\Delta) = \text{Pr}[T_i > t + k\Delta | Y_i(t + k\Delta) \geq y_{i,t+k\Delta}, T_i > t, Y_i(t) \geq y_{i,t}, \ldots]
\]

\[
= \text{Pr}[T_i > t + \Delta | Y_i(t + \Delta) \geq y_{i,t+\Delta}, T_i > t, Y_i(t) \geq y_{i,t}]
\]

\[
\cdot \text{Pr}[T_i > t + 2\Delta | Y_i(t + 2\Delta) \geq y_{i,t+2\Delta}, T_i > t + \Delta, Y_i(t + \Delta) \geq y_{i,t+\Delta}] \cdot \ldots
\]

\[
\cdot \text{Pr}[T_i > t + k\Delta | Y_i(t + k\Delta) \geq y_{i,t+k\Delta}, T_i > t + (k-1)\Delta, Y_i(t + (k-1)\Delta) \geq y_{i,t+(k-1)\Delta}]
\]

\[
= \prod_{j=1}^{k} \text{Pr}[T_i > t + j\Delta | Y_i(t + j\Delta) \geq y_{i,t+j\Delta}, T_i > t + (j-1)\Delta, Y_i(t + (j-1)\Delta) \geq y_{i,t+(j-1)\Delta}, \ldots]
\]

\[
= \prod_{j=1}^{k} \frac{\text{Pr}[Y_i(t + j\Delta) \geq y_{i,t+j\Delta} | T_i > t + (j-1)\Delta, Y_i(t + (j-1)\Delta) \geq y_{i,t+(j-1)\Delta}, \ldots]}{\text{Pr}[Y_i(t + j\Delta) \geq y_{i,t+j\Delta} | T_i > t + (j-1)\Delta, Y_i(t + (j-1)\Delta) \geq y_{i,t+(j-1)\Delta}, \ldots]}
\]

\[
= \prod_{j=1}^{k} \frac{\text{Pr}[y_{i,thresh} \geq Y_i(t + j\Delta) \geq y_{i,t+j\Delta} | y_{i,thresh} \geq Y_i(t + (j-1)\Delta) \geq y_{i,t+(j-1)\Delta}, \ldots]}{\text{Pr}[y_{i,thresh} \geq Y_i(t + j\Delta) \geq y_{i,t+j\Delta} | y_{i,thresh} \geq Y_i(t + (j-1)\Delta) \geq y_{i,t+(j-1)\Delta}, \ldots]}
\]
\[ \prod_{j=1}^{\infty} \frac{\int f(y \mid t + j\Delta)dy}{\int f(y \mid t + j\Delta)dy} \]

where \( \int f(y \mid t + j\Delta)dy \) is the integral of the PDF between the observed degradation index of unit \( i \) and the threshold (illustrated in Figure 4-17, with dash-dot line representing the degradation history of unit \( i \) ); and \( \int_{y_{i,t+j\Delta}}^{\infty} f(y \mid t + j\Delta)dy \) is the integral of the PDF over all possible values equal to or higher than the observed degradation index of unit \( i \) (illustrated in Figure 4-18).
Figure 4-17: Integral of the PDF between the observed degradation value of unit $i$ and the threshold value

Figure 4-18: Integral of the PDF over all probable values equal to or higher than the observed degradation value of unit $i$
4.2.3 Final Target Outputs for ANN Training

The final estimated survival probability of unit \( i \) at interval \( t \) is then the mean of the two survival probability estimates obtained in Sections 4.2.1 and 4.2.2:

\[
S_i(t) = \text{mean} \left[ S_{KM,i}(t), S_{PDF,i}(t) \right]
\]  

(4-39)

The training target vector for historical unit \( i \), denoted here by \( D_i \), consists of the estimated survival probability in the \( h \) successive intervals:

\[
D_i(t) = \begin{bmatrix}
S_i(t + \Delta) \\
S_i(t + 2\Delta) \\
\vdots \\
S_i(t + h\Delta)
\end{bmatrix}
\]  

(4-40)

During training, the input and target vectors of the training sets are repetitively presented to the neural network. The network attempts to produce output values that are as close to the target vectors as possible. After training, when a series of condition values at the current time \( t \) and \( d \) previous time steps

\[
y(t) = \begin{bmatrix}
Y(t) \\
Y(t - \Delta) \\
Y(t - 2\Delta) \\
\vdots \\
Y(t + h\Delta)
\end{bmatrix}
\]  

(4-41)

are fed into the input nodes, the network will produce an output vector

\[
O(t) = \begin{bmatrix}
\hat{S}(t + \Delta) \\
\hat{S}(t + 2\Delta) \\
\vdots \\
\hat{S}(t + h\Delta)
\end{bmatrix}
\]

(4-42)

which can also be plotted as the survival curve estimated at time \( t \). As the next set of input values becomes available, a new updated output vector will be produced, generating a new survival probability curve.
4.2.4 The Proposed FFNN as a Hybrid of Sequence Reproduction Networks and Pattern Classification Networks

Most of the applications of FFNNs tend to fall within the following two categories:

- sequence reproduction, including time series prediction and
- pattern classification, including pattern and sequence recognition and novelty detection.

The way in which the FFNN is used for time series approximation is different from the way it is used for pattern classification. FFNNs which are specialised for reproducing sequences have architectures similar to the example depicted in Figure 4-19. In sequence reproduction, the FFNN is fed a series of sequence elements and generates the following sequence element(s) as output. In pattern classification applications, on the other hand, the network inputs represent feature or attribute entries, while each output corresponds to a class, as shown in Figure 4-20. The hidden units correspond to subclass in this case.
Figure 4-19: Sequence reproduction FFNN

Figure 4-20: Pattern classification FFNN
One of the contributions of the proposed FFNN is that the FFNN combines sequence reproduction networks’ ability to handle sequential input with pattern classification networks’ ability to recognise patterns and output probability representations (Figure 4-21). Sequence reproduction networks are equipped with “memory” to store information over time. Without such memories, each input would be “forgotten” after being mapped to the output. Therefore, sequence reproduction FFNNs have time delays to postpone the forwarding of a node’s activation to another node, and deal with a sequence of past time series elements. At each time step, a single series element is fed into the input, but the network’s forecast takes preceding sequence elements into consideration. While several time delays at the input form a time window, another set of time delays at the hidden layer level duplicates the effect. The input layer of the proposed FFNN adopts the architecture of a sequence reproduction network. This configuration makes the proposed FFNN a powerful tool for handling sequential input. The output layer of the proposed FFNN, on the other hand, adopts the design of a pattern classification network. The pattern of the input sequence is recognised and the contributions for each output class are summed and generated as probabilities. Conventional pattern classification networks usually have a hard-limit transfer function in the output layer to select the maximum of these probabilities, and produce a 1 for that class and a 0 for the other classes. The proposed FFNN, however, uses a saturated linear function instead of a hard-limit function, to produce useful representation of probabilities.
4.3 Summary

This chapter presented a novel approach for overcoming the limitations of the existing prognostic models. The proposed prognostic approach adapts the KM survival analysis technique for modelling suspended histories, the kernel PDE function and conditional reliability for extracting population characteristics from degradation processes, and a FFNN for recognising how degradation is veiled in the non-deterministic changes in CM measurements. Background information on these employed techniques was first given in Section 4.1. The reliability information estimated using the KM estimator and the conditional PDE method were used as target outputs for training the FFNN (see Figure 4-22: Architecture of the proposed prognostic model). The trained FFNN will be able to estimate the future survival curve of an asset, given a series of CM information. The proposed model differs from the existing prognostic model in the literature in several respects. It is a non-parametric approach and is able to extract...
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

population characteristics information from the available historical CM data, both complete and suspended, and integrate this information into prognosis.

Figure 4-23 summarises how data are utilised by the proposed model. The historical event data, which are the times of past failures and suspensions, are used to calculate the overall population survival rate and individual unit survival probability using KM estimator. The available CM histories are used to estimate individual unit reliability using the conditional kernel PDE method. These CM histories are also directly used as training inputs for the ANN. Training targets for the ANN are the mean of the survival probability, estimated using KM estimator and conditional kernel PDE. After training, whenever a series of asset condition data are presented to the network, the network will output an estimate of the asset’s future survival probability.
### Unit-specific Kaplan-Meier Estimation
- Utilises suspended condition monitoring histories
- Models actual survival status of historical assets

<table>
<thead>
<tr>
<th>Condition data</th>
<th>Artificial Neural Network</th>
<th>Survival probability as training targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_t$</td>
<td></td>
<td>$S_{n+1}$ Survival probability in 1&lt;sup&gt;st&lt;/sup&gt; next interval</td>
</tr>
<tr>
<td>$Y_{t-1}$</td>
<td></td>
<td>$S_{n+2}$ Survival probability in 2&lt;sup&gt;nd&lt;/sup&gt; next interval</td>
</tr>
<tr>
<td>$Y_{t-d}$</td>
<td></td>
<td>$S_{n+h}$ Survival probability in $h^{th}$ next interval</td>
</tr>
</tbody>
</table>

### Degradation-based Conditional PDF
- Extracts population characteristics
- Computes conditional reliability based on degradation processes

#### Figure 4-22: Architecture of the proposed prognostic model
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

**Figure 4-23: Utilisation of data by the proposed prognostic model**

**Adapted KM survival estimates for individual failed units**

\[ S_{KR}(t + k\Delta) = \begin{cases} 1, & 0 \leq t + k\Delta < T_i \\ 0, & t + k\Delta \geq T_i \end{cases} \]

\[ \hat{S}(t) = \prod_{i,j \in \mathcal{S}} \left( 1 - \frac{d_{ij}}{n_{ij}} \right) \]

**KM estimation of overall population survival rate**

\[ \hat{S}(t) = \prod_{j=1}^{\infty} f(y \mid t + j\Delta)dy \]

**Degradation-based conditional PDE of survival probability**

\[ S_{PDF}(t + k\Delta) = \prod_{j=1}^{\infty} f(y \mid t + j\Delta)dy \]

**Final estimated survival probability**

\[ S(t) = \text{mean}[S_{KR}(t), S_{PDF}(t)] \]

**ANN training input**

\[ y(t) = [Y(t); Y(t - \Delta); Y(t - 2\Delta); ...; Y(t - d\Delta)] \]

**ANN training target**

\[ O(t) = [\hat{S}(t + \Delta); \hat{S}(t + 2\Delta); ...; \hat{S}(t + h\Delta)] \]
CHAPTER 5 MODEL VALIDATION USING COMPUTER SIMULATED DATA

The previous chapter detailed the development of an intelligent prognostic model which includes suspended condition trending data and population characteristics in prognosis. Multiple sets of asset degradation histories are required to train and test this model. However, this large number of real-life degradation histories is not easy to gather within a short time frame. To overcome this challenge, a computer program for generating synthetic degradation histories was developed. These simulated data will then be used for the performance evaluation and comparison of the proposed model and several other prognostic models.

Section 5.1 describes how defective bearing vibration signatures are modelled and how these modelled signatures can be turned into progressive degradation histories. Section 5.2 shows how these simulated data are used for validating the prognostic model. Finally, the validation results and the corresponding discussions are presented in Section 5.3. The chapter is then summarised and concluded in Section 5.4.
5.1 A Model for the Simulation of Bearing Degradation Histories

Rolling element bearing failure is often the cause of breakdowns in rotating machinery, a common class of machinery in industry. Bearing prognostics research faces a major obstacle: unavailability of failure data. A prognostic model typically requires numerous sets of failure histories for modelling and testing. Unfortunately, even in accelerated run-to-failure tests, it takes weeks or months to fail just one bearing. Often, researchers are required to develop and test their prognostic models on a tight timeline. To resolve this dilemma, a model for simulating bearing degradation histories was developed in this work. This approach provides researchers with much needed data to design and verify their prognostic models before real life data become available.

Braun [145] and McFadden and Smith [145, 146] modelled the vibration response of a bearing with a localised fault as the product of repetitious impulses at the fault frequency, with the bearing load distribution, and the amplitude variation of the transfer function, convolved with an exponential decay due to damping. Although useful for bearing diagnostic purposes, none of these papers considered the modelling of progressively degrading signals. In this work, a vibration waveform – generated by a rolling element bearing under constant radial load with a single point defect – was first modelled using the MATLAB software. This waveform was then repeatedly generated, while the defect severity was increased exponentially with some added random fluctuations.

5.1.1 Modelling Defective Bearing Vibration Signatures

Bearing defects may be categorised as “distributed” or “localised”. Distributed defects include waviness, surface roughness, misaligned races and off-sized rolling elements. Tandon and Choudhury [147] proposed a theoretical model of
the vibration response of radially loaded rolling element bearings to distributed defects. While distributed faults do generate excessive contact forces that can result in premature surface fatigue, they usually do not tend to deteriorate and thus seldom directly result in mechanical failure. It is the localised or point defects such as cracks, pits and spalls on the rolling surfaces that usually lead to failure [148]. This work focuses on the vibration response of rolling element bearings to localised defects.

To describe the waveform generated by a rolling element bearing under constant radial load with a single localised defect, the vibration signature can be expressed as

\[ y(t) = y_d(t) \cdot y_q(t) \cdot y_r(t) \cdot y_e(t) \cdot y_n(t) \]  

(5-1)

where \( y_d(t) \) is a series of impulses at the bearing fault frequency, \( y_q(t) \) is the bearing radial load distribution, \( y_r(t) \) the bearing-induced resonant frequency and \( y_e(t) \) the exponential decay due to damping [146, 149]. The last component, \( y_n(t) \), represents the noise added to corrupt the signal.

Consider a case in which the outer race of a ball bearing has a localised defect (Figure 5-1). Every time a rolling element rolls over the defect, an impulse is produced and this causes the bearing to vibrate at its natural frequency [150]. This vibration response quickly decays because of the damping in the system. As the bearing rotates, this impulse occurs periodically at the outer-race ball pass frequency (BPFO) which is uniquely determined by the defect location. This unique characteristic frequency helps identify the defective component in the machine and the defect location. Therefore, it is this characteristic fault frequency that is of interest in bearing fault diagnostics.
The four characteristic fault frequencies of a ball bearing can be calculated using the following equations [151]:

Ball Pass Frequency, Outer race: \( BPFO(\text{Hz}) = S \left( \frac{N}{2} \right) \left( 1 - \frac{B}{P} \cos \Phi \right) \) (5-2)

Ball Pass Frequency, Inner race: \( BPII(\text{Hz}) = S \left( \frac{N}{2} \right) \left( 1 + \frac{B}{P} \cos \Phi \right) \) (5-3)

Ball Spin Frequency: \( BSF(\text{Hz}) = S \left( \frac{P}{2B} \right) \left( 1 - \frac{B^2}{P^2} \cos^2 \Phi \right) \) (5-4)

Train or Cage Frequency: \( FTF(\text{Hz}) = S \left( \frac{1}{2} \right) \left( 1 - \frac{B}{P} \cos \Phi \right) \) (5-5)

where \( B \) is the ball diameter, \( P \) the pitch diameter, \( N \) the number of balls, \( S \) the shaft rotation speed in Hertz and \( \Phi \) the contact angle in degree.

To describe the components of the model of a faulty bearing’s vibration response, the illustration will be confined to the vibration generated by a single point defect on the inner race of a rolling element bearing under constant radial load.
(i) Repetitious impulses, $y_d(t)$

Each time a rolling element rolls over the inner race defect, an impulse is produced, which can be represented by the impulse function, $\delta(t)$. As the shaft rotates, this impulse occurs periodically at the inner race element passing frequency, $f_i$. The period between the impulses will be denoted as $T_i = \left(\frac{1}{f_i}\right)$.

With amplitude constant $d_0$ denoting the severity of the defect, the series of impulses can be represented mathematically by the equation [152]

$$y_d(t) = d_0 \sum_{j=-\infty}^{\infty} \delta(t - jT_i)$$

(5-6)

Note: the rolling contacts between the ball and the races may be imperfect particularly when a fault exists; therefore, a small random variation in the impulse intervals can be added.

(ii) Non-uniform Load Distribution, $y_q(t)$

Figure 5-2 shows the load distribution around the circumference of a radially loaded rolling element bearing. The angular extent of the load zone is denoted by $\psi_{\text{max}}$. For all $|\psi| < \psi_{\text{max}}$, and the instantaneous load at the contact point of the inner race defect can be determined approximately by using the Stribeck equation [153]; for everywhere else, the instantaneous load at the contact point is zero

$$y_q(t) = \begin{cases} q_0 \left[1 - \left(\frac{1}{2\varepsilon}\right)(1 - \cos \psi)\right]^n & \text{for } |\psi| < \psi_{\text{max}} \\ 0 & \text{elsewhere} \end{cases} \quad (5-7)$$

where $q_0$ and $\varepsilon$ denote the maximum load intensity and the load distribution factor respectively. This amplitude modulation affects the amplitude of defect impulses. The impulse amplitude is assumed to be directly proportional to the instantaneous load on the rolling element when it rolls over the defect.
(iii) Bearing resonant vibration, $v_r(t)$

The impulses excite the natural frequencies of the bearing’s elements and its supporting structures. Under idealised conditions, vibration induced by the bearing at its natural frequency can be represented by a sinusoidal wave

$$ y_r(t) = \sum_{j=-\infty}^{\infty} \sin[2\pi f \left( t - jT_i \right)] $$

(5-8)

where $f_r$ denotes the resonant frequency of the bearing and $T_i$ denotes the period between defect impulses.

(iv) Exponential decay, $v_e(t)$

The resonant vibration is then attenuated exponentially to zero, with a transient duration that depends on the bearing’s damping factor $\alpha$. The decay function can be defined by the equation

$$ y_e(t) = e^{-\alpha_i(t-jT_i)} $$

(5-9)

The product of the previous four model components discussed above is shown in Figure 5-3.

Figure 5-1: Load distribution in a radially loaded bearing
(v) Noise
As the simulated signal will be used for prognostic model training and validation, white Gaussian noise $y_n(t)$ with zero mean and 0.12 standard deviation was added to the signal to simulate real life situations. It is widely known that training data that is assumed sufficiently rich should be generated by a broadband signal [154], such as white Gaussian noise. White Gaussian noise can be readily generated using the “wgn” command in MATLAB.

(vi) Final signal
The final signal $y(t)$ (Figure 5-4) is a product of all the previously defined functions:

$$y(t) = y_d(t) \cdot y_q(t) \cdot y_r(t) \cdot y_e(t) \cdot y_n(t)$$

(5-10)
Based on this method, outer race fault, ball fault and a combination of multiple faults can be simulated. An example is shown here with shaft frequency set at 600rpm. The “bearing” has eleven 7.5mm-diameter balls and a pitch diameter of 34mm. Figure 5-5(a) shows the simulated time domain signal of a bearing with an inner race defect and a ball defect. This signal was converted to frequency domain spectrum using Fast Fourier Transform [Figure 5-5(b)]. The spectrum is dominated by high frequency resonant signals. To separate the bearing fault frequency signal from these dominant signals, the vibration signatures were band-pass filtered. Figure 5-5(c) shows that peaks were detected at 46.8Hz and 67.3 Hz, which closely matched the calculated fault frequencies indicated at the top of Figure 5-5 (BPFI = 67.1 Hz, BSF = 43.1 Hz).

Figure 5-4: Simulated signal of a bearing with an inner race defect after being corrupted by noise
5.1.2 Simulation of Bearing Failure Histories

Now let us consider a scenario which uses the previously simulated data for diagnostic research for prognostic model development. For bearing life prognosis, vibration data are usually collected periodically and data features like RMS are extracted and plotted in order to trend the changes in feature values. These changes are typically taken to indicate the progression of fault severity over time. To simulate progressive degradation data, the defective bearing signal (inner race fault) derived in Section 5.1.1 was repeatedly generated using the “for” looping function in MATLAB. Each repetition represents a measurement recorded at one data collection point. However, just as each real-life degradation data collection would give varied vibration signal with increasing severity, the

![Figure 5-2: Signal of a bearing with multiple faults (a) Time domain plot of raw signal (b) Frequency spectrum of raw signal (c) Fault detection after](image-url)
defect severity parameter $d_0$ of the simulated signal should also be increased at each recording. It has been observed in bearing life tests that bearing degradation signals possess an inherent exponential growth [82]. Therefore, the defect severity parameter $d_0$ for each dataset was programmed to increase exponentially throughout the loop at a randomly selected rate, and to stop at a randomly selected time step (Figure 5-6). One recording or time step could represent, for example, 30 minutes or several weeks in reality.

However, bearing fault progression is usually less deterministic than a smooth exponential curve. For example, when subsurface cracks propagate to the surface and eventually initiate a spall, the spall area increases abruptly whenever material is dislodged from the running surface. Thus, spall growth is discontinuous and some random fluctuations were added to the datasets. These fluctuations also accounted for the effect of real-life disturbances, such as noise, from indirect measurements. The evolution of the defect impulse from one of the datasets is shown in Figure 5-6. This procedure enabled the model to test a prediction model’s capability to adapt to, and capture, the stochastic nature of bearing degradation.
The rest of Equation 5.10 shall remain unchanged because the decay parameter, natural frequency, noise, and duty parameters of the bearing being monitored remain the same. For illustration, one of every 10 recordings of one of the datasets is plotted in Figure 5-7. It can be seen that defect severity increased exponentially from the first to the last recording.

Figure 5-3: Evolution of defect impulse in dataset No.1 – exponential growth with random fluctuations
Having these progressive bearing degradation histories, we can now extract as many features as desired to be used as the input data for a prognostic model. Some of the extracted features are shown in Figure 5-8. RMS was fairly stable but might not be a very sensitive indicator of an incipient fault. Kurtosis provided a strong indication of a developing problem in the intermediate stages.

Figure 5-7: Defect severity increasing exponentially (only one of every 10 datasets generated are plotted here)
(at i≈40-70) but did not represent the defect progression well near the end of life. These observations were consistent with the feature analysis reported in other bearing prognostic research [76, 155] and showed that features extracted from these simulated signals have the same characteristics as those extracted from real life test signals. Entropy estimation [156] and peak value were found to provide stable and consistent indication throughout the defect progression.

Figure 5-8: Sample features extracted from simulated data
5.2 Procedure of Model Validation Using Simulated Signals

For the validation of the proposed prognostic model, 35 degradation histories were simulated. Each history had defect impulses that increased at different rates with different random fluctuations. Defect impulses of Datasets No.1 – No.3 are plotted as examples in Figure 5-9.

Thirty of the 35 datasets were assigned for model training and the remaining five for model testing (Table 5-1). Suspensions were imposed on 1/3 of the training datasets at different time steps. Four time steps were grouped as one interval.

![Figure 5-4: Examples of the computer-simulated degradation datasets](image)
### Table 5-1: Datasets simulated for model training and testing

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time of failure/suspension (time step)</th>
<th>Training Datasets that end in failure</th>
<th>Training Datasets that end in suspension</th>
<th>Datasets used for model testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 5.2.1 Processing of Simulated Signals

Various features were calculated from the 35 sets of simulated signals. The features selected for representing defect progression are peak, RMS and entropy estimation. Even though these simple time domain features are not usually able to distinguish between the influence of defect progression in the signal from the influence of operating conditions [85], it was found that they sufficed in these...
simulation tests (with a constant level of simulated background noise and operating conditions).

Past research has shown that each signal feature is only effective for certain defects at certain stages [155]. According to Huang et al. [83], the self-organising map (SOM) mean quantisation error (MQE) takes advantage of mutual information from multiple features and is therefore a good degradation indicator. SOMs [157], as discussed in Section 4.1.1, are ANNs that self-organise when learning. A SOM was used in this work to map the high-dimensional input data (three trending features) onto a low-dimensional output space (a single degradation indicator).

Firstly, feature values calculated from initial healthy state were used to train the SOM by moving its weights to mark the “healthy state” region in the map. Figure 5-10 shows the organisation of the weights of a 3-dimensional SOM before and after training. The weight vectors, shown with crosses, were randomly initialised. The three axes indicate the coordinates of the weights in three directions. During training, defect-free or healthy bearing data were inputted into the SOM as training samples for marking the “healthy state”. The organisation of these training samples is shown in Figure 5-11 as circles. The map went through all the weight vectors and calculated the distance from each weight to the chosen training sample. The weight with the shortest distance was the best matching unit and would be updated to become more like the selected training sample.
After training, degradation data of each bearing can be fed into the map as test data. Figure 5-12 shows a sample trajectory of a set of bearing degradation data. The three coordinate axes indicate the location of the weights and data points. When the defect impulses were very small and masked by noise, the defect data (shown with circles) were very close to the healthy state, indicated by crosses. As the defect severity increased, the deviation of the input from the normal state became increasingly larger.
The deviations of these input data vectors from the closest weights in the “healthy state” region were calculated. These deviations are called the Quantisation Errors, defined as

$$QE(t) = x_j(t) - w_{j\ast}(t)$$  \hspace{1cm} (5-11)

where $x_j$ is the map input vector with $j$ dimensions, $w_{j\ast}$ is the closest weight. Similar to the MQE proposed by Huang et al. [83], the QE can serve as a defect severity indicator which trends the deviation of the data feature values from the healthy state values. Figure 5-12 depicts the QE curve derived from degradation dataset No.1.

Figure 5-5: A sample trajectory of a 3-dimensional bearing degradation dataset in a SOM
In this way, the various data features were combined into a one-dimensional feature by the SOM. This procedure reduces the number of input nodes and parameters of the FFNN used in the prognostic model and, in turn, reduces the network training time. The prognostic model will take in the one-dimensional degradation indices and estimate the future survival probabilities of the bearing of interest.

Figure 5-13 summarises how the simulated data are processed and used for prognostic model testing. Features are first calculated from the simulated data and then combined into a single degradation indicator using the SOM technique. This degradation indicator is fed into the prognostic model for training and prediction.

Figure 5-6: Trending defect severity by using SOM Quantisation Error as a degradation indicator
Figure 5-13: The high-dimensional feature matrix was reduced to a one-dimensional input vector and fed into the prognostics model.
5.2.2 Design of Performance Evaluation

A significant methodological issue is the performance evaluation of the prognostic model. As the prediction output of the proposed model is survival probabilities, the exact predicted failure times are not represented. Still, there is a well-defined goal: the accurate prediction of individual prognosis. Each of the five prediction results will be analysed individually. For evaluation purposes, the predicted failure time was identified merely by noting the first output unit that gave a survival probability of less than 0.5. This instance represents the point in time during training that the neural network’s target output changes from 1 to 0 immediately after failure occurs. Survival probability of 0.5 was also chosen because it is the value used in KM estimation to determine the median survival time. It is also the probability value for determining the commonly used mean time to failure (MTTF) in Weibull distribution – one of the models the proposed model is compared against in the following chapters. The failure times determined in this way by the proposed model were compared with the actual failure times.

(i) Model Comparison

One of the main objectives of this research is to develop a method for including suspended CM histories in asset health prognosis. It was suggested in Chapter 3 that the future health or remaining useful life of assets would be underestimated if historical suspension times were modelled as historical failure times. On the other hand, excluding historical suspended trending data from modelling would worsen the problem of data unavailability. To validate the assertion that including suspended condition trending data improves the prognostic capability of a model, the performance of the proposed model was compared to those of:

- an FFNN with the same structure and training function as the proposed model, but which treated suspension times as failure times (Model A), and
- an FFNN with the same structure and training function as the proposed model, but which excluded suspended histories from training examples (Model B).
The training target vectors for the complete histories were the same for all three models. As for the suspended histories, training target vectors for the proposed model were computed as discussed in the previous chapter. Model A, on the other hand, was trained based on the false assumption that all suspension times were failure times. In other words, the training target vectors for suspended datasets were computed the same way as those for complete failure datasets. Model B was trained using only the complete failure histories. The simulation test consisted of three assessments (Table 5-2). In Assessment I, all 20 complete datasets and 10 suspended ones were made available for model training. In Assessment II, only 10 complete datasets and 10 suspended ones were used. In the last assessment, only 1 complete and 10 suspended datasets were used.

Table 5-2: Datasets used for training in the simulation assessments

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Failure</td>
</tr>
<tr>
<td>I</td>
<td>20</td>
</tr>
<tr>
<td>II</td>
<td>10</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
</tr>
</tbody>
</table>

Another main objective of this research is to investigate whether including population characteristics in training improves the prognostics performance of a model. Therefore, the proposed model was also compared with

- a recurrent neural network (RNN) that approached machine health prognosis as a time series prediction problem (Model C).

The RNN was selected as one of the control models because it is the most commonly used AI prognostic models found in the literature [1, 7-9]. These
time series prediction models generally forecast the condition value at the next immediate time step(s) and require a predetermined failure threshold.

Lastly, to evaluate the performance of the proposed prognostic model against classical reliability models, the proposed model was also compared with

- a traditional Weibull model (Model D)

The classical Weibull distribution was used to model the simulated failure times and estimate the mean lifetime of the units. As mentioned, Weibull model is among the most popular and commonly used failure models due to its flexible shape and ability to model a wide range of failure rates. The failure PDF of the Weibull distribution is defined as

\[ f(t) = \frac{\beta}{\eta} \left(\frac{t - \gamma}{\eta}\right)^{\beta-1} e^{\left(\frac{t - \gamma}{\eta}\right)^\beta} \]

where \( \beta \) is the shape or slope parameter, \( \eta \) is the scale parameter, and \( \gamma \) is the location constant which indicates the minimum time when a failure starts. The two-parameter Weibull PDF is obtained by setting \( \gamma = 0 \). The shape and scale parameters were determined using maximum likelihood estimation (MLE). MLE involves developing a likelihood function based on the available data and finding the values of the parameter estimates that maximize the likelihood function. The mean of the Weibull PDF, \( \bar{T} \), which is the MTTF in this case, is given by

\[ \bar{T} = \eta \cdot \Gamma \left( \frac{1}{\beta} + 1 \right) \]

where \( \Gamma \) is the well known gamma function.
(ii) Performance Metrics

To compare the performance of the above models, appropriate performance metrics need to be developed. There is yet to be a uniform set of benchmark criteria for assessing the technical performance of prognostic models. Vachtsevanos et al. [11] suggested that quantitative measures of the prediction performance should consider both the closeness of the predicted value to the actual value, as well as the width of the confidence interval. In the candidate’s view, another important performance measure is the length of prediction horizon; that is, it is important to consider how far into the future a prognostic model can predict. For example, a prognostic model that can predict with 80 percent accuracy one week before failure delivers a higher performance than a model that can predict with the same accuracy an hour before failure.

Vachtsevanos et al. [158] stated that it is also important to consider the relative position of the predicted time-to-failure along the time axis with respect to the occurrence of the actual failure. They provided three example scenarios. In scenario (a), the failure occurs later than predicted; in scenario (b), failure occurs within the predicted failure interval; and in scenario (c), the failure occurs much earlier than predicted. Scenario (b) is preferred over scenarios (a) and (c). However, it is scenario (c) that can cause catastrophic consequences and should be avoided as much as possible. They concluded that accuracy and precision metrics may need to be augmented with severity requirements as to the relative timing of the predicted failure time with reference to the actual failure occurrence.

In this work, a penalty function was defined. The function considers the mean prediction accuracy and the prediction horizon of a prognostic model


given by

\[
p(y) = \frac{1}{c} \sum_{j=1}^{c} \left[ p_g(y_j) \right] + p_h(y)
\]

(5-14)

where \( c \) is the number of test-sets.
The prediction accuracy function \( p_g \) measures the discrepancy between the actual failure time \( T \) and the predicted failure time \( \hat{T} \) in each test-set

\[
p_g(y) = \begin{cases} 
\alpha(T - \hat{T}), & \hat{T} < T \\
0, & \hat{T} = T \\
\beta(\hat{T} - T), & \hat{T} > T 
\end{cases}
\]

(5-15)

where \( \alpha \) and \( \beta \) are penalty parameters of underestimation and overestimation in failure time prediction, respectively, and \( \alpha < \beta \), since overestimation is worse than underestimation in failure time prediction.

The prediction horizon function \( p_h \) subjects penalty to exponential decay as the length of horizon increases

\[
p_h(y) = e^{-\lambda h}
\]

(5-16)

where \( \lambda \) is the decay constant.

In this simulation test, the values of \( \alpha \), \( \beta \) and \( \lambda \) were arbitrarily assigned with 0.1, 0.15 and 0.2 respectively for evaluation purposes.

5.2.3 Determining the Structure of the Prognostic Models

ANNs can be constructed in many ways. This section discusses the structure of the FFNNs and the RNN used for this simulation test.

(i) Feed-Forward Neural Networks

A number of practical considerations were taken before choosing the FFNN structure.
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

Firstly, the number of neurons in the input and output layers was determined by the nature of the problem. In this test, eight output nodes and 11 input nodes were chosen for each of the FFNNs so that the networks could predict eight time steps ahead based on the “current” asset condition value and 10 delayed values. This number of input nodes was found to be optimal for the simulated data.

The next thing to be determined was the number of hidden layers and the number of hidden neurons. There is currently no theoretical reason to use FFNNs with any more than two hidden layers. For most practical problems, only one hidden layer is required. The results of using different numbers of hidden layers are summarised in Table 5-3.

<table>
<thead>
<tr>
<th>Number of Hidden Layers</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Only capable of representing linear separable functions or decisions.</td>
</tr>
<tr>
<td>1</td>
<td>Capable of approximating arbitrarily any functions which contains a continuous mapping from one finite space to another.</td>
</tr>
<tr>
<td>2</td>
<td>Capable of representing an arbitrary decision boundary to arbitrary accuracy with rational activation functions and able to approximate any smooth mapping to any accuracy.</td>
</tr>
</tbody>
</table>

In our case, which involves continuous mapping between finite spaces, one hidden layer was found to be sufficient. Deciding the number of hidden neurons is also a very important part of deciding the overall network architecture. Though these layers do not directly interact with the external environment, these layers have a significant influence on the final output. Having too few hidden neurons prevents the network from correctly mapping inputs from complicated datasets to outputs, while having too many causes over-fitting of the data and
impedes generalisation. A general rule-of-thumb method for determining the correct number of hidden neurons is that the number of hidden neurons should be two-third (2/3) of the number of input nodes, plus the number of output nodes. This rule was used as a starting point. Ultimately, the selection of the architecture of ANNs comes down to trial and error [160]. The “forward selection” trial-and-error method was used. The FFNNs were first trained and tested with a small number of hidden neurons – four in this case. The results were analysed and recorded. The process was repeated with an increased number of hidden neurons so long as the overall results of the training and testing improved (Figure 5-14). The optimal number of hidden nodes found for this simulation test was 11. Other than selecting an optimal number of hidden nodes, the “early stopping” method was also used to prevent over-fitting. When choosing the number of training epochs, data were divided into three subsets: the train-set, the validation set and the test-set. The error in the validation set was monitored during the training process. Once the validation error increased for around 25 iterations (instead of continuing to decrease), the training was stopped and the weights and biases at the minimum validation error were returned.
Figure 5-14: Selecting the number of hidden neurons with forward selection

1. Start
2. Select a small number of hidden neurons.
3. Train the neural network and evaluate the network performance.
4. Is the performance better than that with the previous choice of number of neurons?
   - Yes: Stop. Select the previous choice of number of neurons.
   - No: Add a neuron.
The transfer function selected for the hidden layer was the tangent sigmoid function. It calculates each hidden node’s output from the node’s net input, $\nu$, as follows:

$$f(\nu^{(l-1)}) = \frac{2}{1 + e^{-2\nu^{(l-1)}}} - 1$$

(5-17)

where $i$ is the index of nodes in layer $l$. This algorithm is mathematically equivalent to the hyperbolic tangent function with negligible numerical differences. The tangent sigmoid function runs faster in MATLAB implementation and is therefore a good trade-off for neural network applications, where computation speed is important and the exact shape of the transfer function is not. The function is illustrated in Figure 5-15.

![Figure 5-15: Tangent sigmoid transfer function](image)

The transfer function selected for the output layer is the saturated linear function. Using this function, the output node’s activation is directly proportional to the node’s net input, $\nu$, and truncated into the interval [0, 1]. The function is illustrated in Figure 5-16. The use of such transfer function in the network output layer plays an important role in allowing the outputs to be given a probabilistic interpretation.
Figure 5-7: Saturated linear transfer function

Other factors that may influence a network’s performance include the selection of training functions. Generally, there are no detailed criteria to be followed while implementing a network. Most of the time, both experience and trial-and-error method are required [132]. The training function used in this work was the back-propagation algorithm with momentum which provides fast convergence [161]. Standard back-propagation is a gradient descent (also called steepest descent) algorithm, in which the network weights are moved along the negative of the gradient of the performance function. The term “back-propagation” refers to the manner in which the gradient is computed for non-linear multilayer networks. The network weights and biases are updated in the direction in which the performance function decreases most rapidly - the negative of the gradient. Properly trained back-propagation networks tend to give reasonable answers when presented with inputs that they have never seen. Typically, a new input leads to an output similar to the target output for input vectors used in training, which are similar to the new input being presented. This generalisation property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output
pairs. By adding momentum to the learning algorithm, each new weight vector, $w_k$, is adjusted in the following fashion:

$$\partial w_k = \mu \partial w_{k-1} + \alpha (1 - \mu) \frac{\partial E}{\partial w_k}$$  \hspace{1cm} (5-18)

where $\mu$ is the momentum constant, $\alpha$ is the learning rate and $E$ is the network performance. Back propagation is used to calculate the derivatives of performance, $\partial E$, with respect to the weight and bias variables. The gradient is calculated by summing the gradients calculated at each training example, and the weights and biases are updated after all training examples have been presented. The learning procedure requires that the change in weight is proportional to $\frac{\partial E}{\partial w_k}$. The constant of proportionality is the learning rate $\alpha$. The performance of the algorithm is sensitive to the proper setting of the learning rate. If the learning rate is set too high, the algorithm may oscillate and become unstable. On the other hand, if the learning rate is too low, the algorithm can take too long to converge. In fact, the optimal learning rate changes during the training process, as the algorithm moves across the performance surface. For practical purposes, a learning rate should be as large as possible, without leading to oscillation. The learning rate for this test was set to 0.02. The addition of a momentum term also avoids oscillation at large by making the change in weight dependent on the past weight change. Momentum allows the network to respond not only to the local gradient, but also to recent trends in the error surface. Functioning like a low-pass filter, momentum enables the network to ignore small features in the error surface and slide through local minimums without getting stuck. The magnitude of the effect that the last weight change is allowed to have is mediated by the momentum constant, $\mu$, which can be any value in the range $[0, 1]$. A momentum constant of 0 means weight change depends solely on the gradient, whereas a momentum constant of 1 implies that the new weight change is set to equal the last weight change, and gradient is ignored. The momentum constant was set to be 0.9 for this test.
(ii) Recurrent Neural Network

The RNN selected for comparison here was an Elman network [10]. An Elman network had a feedback connection from the hidden layer to the input layer and hence can detect time-varying patterns. This recurrent connection has a delay that stores information from the previous time steps for future reference. The Elman network used in this work had hidden nodes that used a tangent sigmoid transfer function and output nodes that used a linear transfer function. This combination allowed a two-layer network to approximate any function (with a finite number of discontinuities) with arbitrary accuracy [2]. Tangent sigmoid transfer function has been introduced in the previous subsection. Linear transfer function sets the node output to vary in direct proportion to the node’s net input, as illustrated in Figure 5-17.

![Figure 5-17: Pure linear transfer function](image)

The network had nine hidden nodes and predicted three steps ahead. This structure was selected based on the best trade-off between structure complexity, prediction horizon length and prediction accuracy obtained through a post-training regression analysis of validation. Three parameters were used for the regression result evaluation: the slope and the y-intercept of the best linear
regression relating targets to outputs, as well as the correlation coefficient (R-value) between the outputs and targets. If we had a perfect fit and a perfect correlation, the slope would be 1, the y-intercept would be 0, and the R-value – which is a measure of how well the variation in the output is explained by the targets – would be 1. The three-step-ahead RNN with nine hidden nodes produced an R-value of 0.979, which was very close to 1 and therefore indicated a reasonably good fit. Figure 5-18 illustrates the graphical output from the regression analysis.

The network outputs are plotted against the targets as open circles. The best linear fit is indicated by a dashed line whereas the perfect fit (output equal to targets) is indicated by the solid line. The RNN used the Levenberg-Marquardt back propagation training function, which updates weight and bias values
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

according to Levenberg-Marquardt optimisation. The optimal number of training epochs was determined by using the early stopping method. A condition indicator value of 1.10 was selected as the failure threshold of all the generated datasets.

5.3 Simulation Results and Discussion

The survival function for the entire training datasets estimated using the standard KM estimator is depicted in Figure 5-19.
The five sets of prediction results from each of the five models were analysed individually. Subsections 5.3.1 - 5.3.5 discuss the prognosis results of the proposed model and Models A-D respectively. As there are five models to be discussed and each model has 15 sets of prediction results (5 test-sets \( \times \) 3 assessments), only the first test-set in Assessment I will be discussed in detail. The other results will be summarised and compared in Subsection 5.3.6.

**Figure 5-19: The standard KM Estimation of survival probability for the 30 training datasets**
5.3.1 Results of Assessment I Produced by the Proposed Model

Figure 5-20 shows one of the sample input datasets (Dataset No.35 in which the actual failure time was simulated at $t=69$), and Table 5-4 shows the corresponding prediction results produced by the proposed model.
Table 5-4: Prediction output of the proposed model for simulated dataset No. 35 in Assessment I

<table>
<thead>
<tr>
<th>t=11</th>
<th>t=12</th>
<th>t=13</th>
<th>t=14</th>
<th>t=15</th>
<th>t=16</th>
<th>t=17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=18</th>
<th>t=19</th>
<th>t=20</th>
<th>t=21</th>
<th>t=22</th>
<th>t=23</th>
<th>t=24</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=32</th>
<th>t=33</th>
<th>t=34</th>
<th>t=35</th>
<th>t=36</th>
<th>t=37</th>
<th>t=38</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=46</th>
<th>t=47</th>
<th>t=48</th>
<th>t=49</th>
<th>t=50</th>
<th>t=51</th>
<th>t=52</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.97</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.96</td>
<td>0.92</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>1</td>
<td>0.99</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>1</td>
<td>0.98</td>
<td>0.96</td>
<td>0.95</td>
<td>0.92</td>
<td>0.84</td>
<td>0.87</td>
</tr>
<tr>
<td>1</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.91</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>1</td>
<td>0.95</td>
<td>0.92</td>
<td>0.93</td>
<td>0.89</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>1</td>
<td>0.94</td>
<td>0.89</td>
<td>0.90</td>
<td>0.92</td>
<td>0.88</td>
<td>0.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=60</th>
<th>t=61</th>
<th>t=62</th>
<th>t=63</th>
<th>t=64</th>
<th>t=65</th>
<th>t=66</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.88</td>
<td>0.83</td>
<td>0.84</td>
<td>0.86</td>
<td>0.73</td>
<td>0.66</td>
<td>0.59</td>
</tr>
<tr>
<td>0.84</td>
<td>0.78</td>
<td>0.81</td>
<td>0.78</td>
<td>0.65</td>
<td>0.57</td>
<td>0.47</td>
</tr>
<tr>
<td>0.82</td>
<td>0.77</td>
<td>0.78</td>
<td>0.73</td>
<td>0.60</td>
<td>0.50</td>
<td>0.38</td>
</tr>
<tr>
<td>0.78</td>
<td>0.71</td>
<td>0.70</td>
<td>0.63</td>
<td>0.52</td>
<td>0.42</td>
<td>0.28</td>
</tr>
<tr>
<td>0.71</td>
<td>0.61</td>
<td>0.61</td>
<td>0.58</td>
<td>0.43</td>
<td>0.33</td>
<td>0.18</td>
</tr>
<tr>
<td>0.64</td>
<td>0.56</td>
<td>0.52</td>
<td>0.47</td>
<td>0.34</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>0.60</td>
<td>0.50</td>
<td>0.43</td>
<td>0.40</td>
<td>0.26</td>
<td>0.14</td>
<td>0.00</td>
</tr>
<tr>
<td>0.55</td>
<td>0.41</td>
<td>0.35</td>
<td>0.32</td>
<td>0.16</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Failed at t=69
The prediction output at each time step, $O(t)$, is arranged vertically, with its 1st row value representing the probability of the unit surviving the immediate next interval, 2nd row value representing the probability of the unit surviving the 2nd following interval, and so on up to the 8th row. It was observed that the predicted survival probabilities closely matched the actual degradation trend. The survival probability was unity (1) initially, and began to drop at the 47th time step as the unit degraded considerably. The predicted survival probability first fell below 0.5 at the 8th row of $t=61$ (“0.41”, highlighted in Table 5-4). This fall at the 8th row means that the unit was expected to fail at the 8th following time step; that is, $t=69$. At $t=62$. The first prediction output that displayed a probability of below 0.5 was at the 7th row, which means the unit was expected to fail at the 7th following time step; that is, $t=69$. This implication can be observed consistently in the following prediction output. The actual simulated failure was indeed at the 69th time step. Tests using the other four datasets produced similarly promising results.

As mentioned, the probability values of the output nodes can be combined to form an estimated survival curve for an individual bearing. Figure 5-21 depicts the survival curve represented by the output values at $t=64$. 
Note that the survival probability curve for Dataset 35 at this time was indeed different from the cumulative survival probability curve for the entire training datasets (Figure 5-19) at that time. This observation indicates that the proposed model produces customised survival prediction for individual bearings by adapting the survival probabilities with more specific health condition data, rather than only fitting a statistical failure distribution (e.g. Weibull distribution) to the entire set of failure event data.

5.3.2 Results of Assessment I Produced by Model A

Model A underestimated the failure times in all five test-sets. Table 5-5 shows the prediction results for Dataset No.35. The probability value fell below 0.5 at
\( t = 54 \) in the 8th row, meaning that the bearing was expected to fail around 8 time steps later at \( t = 62 \). However, the failure was actually simulated to occur at \( t = 69 \). This observation indicates that if suspensions are not modelled correctly and are simply treated as failures, the remaining useful life of assets will be underestimated. The survival probability predicted by Model A at the remaining time steps of dataset No. 35 (and other datasets) was similarly underestimated.
Table 5-5: Predicted survival probabilities produced by Model A for the test-set No.1 in Assessment I

<table>
<thead>
<tr>
<th>t=2</th>
<th>t=3</th>
<th>t=4</th>
<th>t=5</th>
<th>t=6</th>
<th>t=7</th>
<th>t=8</th>
<th>t=9</th>
<th>t=10</th>
<th>t=11</th>
<th>t=12</th>
<th>t=13</th>
<th>t=14</th>
<th>t=15</th>
<th>t=16</th>
<th>t=17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Surv. probability in 1st following interval \( \hat{S}_{t+\Delta t} \)
\( \hat{S}_{t+2\Delta t} \)
\( \hat{S}_{t+3\Delta t} \)
\( \hat{S}_{t+4\Delta t} \)
\( \hat{S}_{t+5\Delta t} \)
\( \hat{S}_{t+6\Delta t} \)
\( \hat{S}_{t+7\Delta t} \)
\( \hat{S}_{t+8\Delta t} \)

- 144 -
5.3.3 Results of Assessment I Produced by Model B

Table 5-6 shows the Dataset No.35 prediction results produced by Model B in Assessment I. The predicted survival probability value first dropped to below 0.5 in column $t = 59$ in the 8\textsuperscript{th} row, meaning that the monitored bearing was forecasted to fail around $t = 67$, which was 2 days before the actual failure. The predicted failure time never deviated more than two time steps from the actual failure time. The reasonable prediction accuracy might be due to the relatively large number of failure datasets available for training in this assessment. Even though this model did not utilise suspended data for training, the 20 failure datasets were sufficient learning examples for the network. However, the model did not perform well in other assessments where less or no failure datasets were available for model training.
Table 5-6: Predicted survival probabilities produced by Model B for dataset No.35 in Assessment I

<table>
<thead>
<tr>
<th>t=11</th>
<th>t=12</th>
<th>t=13</th>
<th>t=14</th>
<th>t=15</th>
<th>t=16</th>
<th>t=17</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=18</th>
<th>t=19</th>
<th>t=20</th>
<th>t=21</th>
<th>t=22</th>
<th>t=23</th>
<th>t=24</th>
<th>t=25</th>
<th>t=26</th>
<th>t=27</th>
<th>t=28</th>
<th>t=29</th>
<th>t=30</th>
<th>t=31</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=32</th>
<th>t=33</th>
<th>t=34</th>
<th>t=35</th>
<th>t=36</th>
<th>t=37</th>
<th>t=38</th>
<th>t=39</th>
<th>t=40</th>
<th>t=41</th>
<th>t=42</th>
<th>t=43</th>
<th>t=44</th>
<th>t=45</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=46</th>
<th>t=47</th>
<th>t=48</th>
<th>t=49</th>
<th>t=50</th>
<th>t=51</th>
<th>t=52</th>
<th>t=53</th>
<th>t=54</th>
<th>t=55</th>
<th>t=56</th>
<th>t=57</th>
<th>t=58</th>
<th>t=59</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.96</td>
<td>0.93</td>
<td>0.93</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
<td>0.94</td>
<td>0.97</td>
<td>0.99</td>
<td>1</td>
<td>1</td>
<td>0.98</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.97</td>
<td>0.93</td>
<td>0.91</td>
<td>0.92</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.88</td>
<td>0.84</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.98</td>
<td>0.96</td>
<td>0.97</td>
<td>0.99</td>
<td>0.93</td>
<td>0.92</td>
<td>0.89</td>
<td>0.89</td>
<td>0.85</td>
<td>0.81</td>
<td>0.80</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.89</td>
<td>0.88</td>
<td>0.92</td>
<td>0.89</td>
<td>0.86</td>
<td>0.78</td>
<td>0.73</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.88</td>
<td>0.86</td>
<td>0.91</td>
<td>0.87</td>
<td>0.81</td>
<td>0.71</td>
<td>0.68</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.97</td>
<td>0.82</td>
<td>0.80</td>
<td>0.83</td>
<td>0.77</td>
<td>0.71</td>
<td>0.64</td>
<td>0.62</td>
<td>0.51</td>
</tr>
<tr>
<td>1</td>
<td>0.99</td>
<td>0.94</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.81</td>
<td>0.76</td>
<td>0.75</td>
<td>0.68</td>
<td>0.65</td>
<td>0.61</td>
<td>0.57</td>
<td>0.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=60</th>
<th>t=61</th>
<th>t=62</th>
<th>t=63</th>
<th>t=64</th>
<th>t=65</th>
<th>t=66</th>
<th>t=67</th>
<th>t=68</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>0.86</td>
<td>0.73</td>
<td>0.68</td>
<td>0.66</td>
<td>0.59</td>
<td>0.52</td>
<td>0.45</td>
<td>0.36</td>
</tr>
<tr>
<td>0.82</td>
<td>0.78</td>
<td>0.65</td>
<td>0.63</td>
<td>0.57</td>
<td>0.48</td>
<td>0.42</td>
<td>0.27</td>
<td>0.21</td>
</tr>
<tr>
<td>0.79</td>
<td>0.73</td>
<td>0.61</td>
<td>0.54</td>
<td>0.51</td>
<td>0.38</td>
<td>0.25</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>0.71</td>
<td>0.63</td>
<td>0.53</td>
<td>0.43</td>
<td>0.42</td>
<td>0.28</td>
<td>0.18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.61</td>
<td>0.57</td>
<td>0.42</td>
<td>0.38</td>
<td>0.31</td>
<td>0.15</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.52</td>
<td>0.47</td>
<td>0.31</td>
<td>0.22</td>
<td>0.19</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.42</td>
<td>0.37</td>
<td>0.22</td>
<td>0.11</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.33</td>
<td>0.29</td>
<td>0.13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Failed at t=69
5.3.4 Results of Assessment I Produced by Model C

Figure 5-22 presents the prediction results of Model C, which was an Elman network that used the condition index to represent the unit health in prognostic applications. The Elman network could predict the degradation index value fairly well, especially when the simulated data did not have many fluctuations. However, this sort of time series prediction model relies on a predetermined failure threshold, which is not always consistent for every dataset. The prediction horizon was also limited to just three time steps ahead – even for the simple simulated degradation trend.

Figure 5-22: The RNN prediction results of test-set No.1 in Assessment I
5.3.5 Results of Assessment I Produced by Model D

For this simulation test, the estimated values for the Weibull shape parameter $\beta$ and the scale parameter $\eta$ are 8.62 and 77.69, respectively. The Weibull plot produced by Model D (traditional Weibull model) is shown in Figure 5-23. The estimated mean for the simulated failures was 73.4. In other words, the average time that each of the bearings in the population was expected to operate before failure was $t = 73.4$. The model did not provide a customised health forecast for each individual bearing. All the simulated failures in the test-sets occurred at least four time steps earlier or later than the predicted failure time.

Figure 5-23: Weibull plot by Model D
5.3.6 Overall Model Comparison for All Assessments

The penalty points of the proposed model and Models A, B and C are presented in Table 5-7. The values of $\alpha$, $\beta$ and $\lambda$ were assigned with 0.1, 0.15 and 0.2, respectively.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Proposed</th>
<th>A (models suspensions as failures)</th>
<th>B (excludes suspensions from training)</th>
<th>C (one-step-ahead time series prediction)</th>
<th>D (traditional Weibull model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (20F,10S)</td>
<td>0.868</td>
<td>1.568</td>
<td>1.101</td>
<td>1.119</td>
<td>3.167</td>
</tr>
<tr>
<td>II (10F,10S)</td>
<td>1.035</td>
<td>1.901</td>
<td>1.035</td>
<td>1.119</td>
<td>10.350</td>
</tr>
<tr>
<td>III (1F,10S)</td>
<td>0.785</td>
<td>2.001</td>
<td>8.168</td>
<td>1.152</td>
<td>25.95</td>
</tr>
</tbody>
</table>

The proposed model had the lowest penalty point in all three assessments. Model A was greatly penalised as it underestimated the time to failures. The performance of Model B was quite good in Assessments I and II where there were a reasonable number of complete failure datasets available for training. However, when there were only suspended data available for training, Model B was totally incapable of performing prediction. Model C received very high penalty points due to its short prediction horizon. Also, the time that the predicted degradation index crossed the predetermined threshold did not match the failure time. Model D, which was based solely on reliability data, was penalised considerably, because the randomly simulated time-to-failures did not match the expected failure time it estimated. The Weibull model only provided an average failure time for all the five test units and did not account for their condition data. Nonetheless, the prediction horizon penalty points this model received were negligible since its prediction horizon was long; this was because the MTTF could be calculated even before operation started. Similar to Model
B, Model D was also incapable of estimating survival when there were no failure data available.

The results support the assertion that the proposed model gives better prediction performance than do the other four models. It is, however, acknowledged that the choice of penalty function parameters $\alpha, \beta$ and $\lambda$ could introduce bias into the comparative analysis. Nonetheless, no benchmark criteria have been developed for assessing or comparing the technical performance of machinery prognostic models. The proposed performance metrics set the stage for the comparison of reliability and prognostic models and are, therefore, considered appropriate for evaluation purposes in this work.

5.4 Summary

This chapter first detailed the development of a model for simulating bearing degradation histories. The simulation model is able to generate a large number of datasets to provide researchers with much needed data for developing and testing their prognostic models before real life data become available. Random fluctuations were added into the simulated data in order to test a prognostic model’s capability of adapting to, and capturing, the stochastic nature of asset degradation. A comparison of features extracted from the simulated data with those from real data suggests that the proposed simulation model is a viable research tool for bearing prognostics. For this study, multiple sets of simulated degradation histories were generated and suspensions were imposed on a portion of the datasets. These data were used to train and test the proposed prognostic model, as well as to compare the proposed model with four control models. The results indicate that the proposed model performed better than: (a) similar models that neglected suspended data; (b) an RNN time series prediction model; and (c) the traditional Weibull model. The promising results suggest that the proposed concept is worthy of further study. In
the following chapter, real life bearing degradation data gathered from industry will be used to further evaluate the performance of all five models.
CHAPTER 6  MODEL VALIDATION
THROUGH INDUSTRY CASE STUDY

Since one of the main objectives of this work is to develop a practicable prognostic model, the proposed model was evaluated through application to real life data. The condition and event data of centrifugal pumps at Irving Pulp and Paper\(^1\) were used for the training, testing and comparison of the proposed model, as well as the other four models described in the previous chapter. Section 6.1 briefly introduces the pulp and paper mill and the operation of the centrifugal pumps used for this study. Section 6.2 describes the data obtained, whereas Section 6.3 discusses the challenges of applying the proposed model to real life prognosis. The assessment results of each model are presented in Section 6.4.

6.1 Pump Condition Monitoring at Irving Pulp and Paper Mill

Pulp and paper is one of the capital-intensive, low-margin, highly competitive commodity industries that use CBM extensively. Examples of such companies include the Swedish company Stora Enso that employs continuous monitoring due

\(^1\) Data courtesy of the Center for Optimization & Reliability Engineering (C-MORE), University of Toronto, and Irving Pulp and Paper, Saint John, Canada.
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data
to the high costs of production loss and the secondary costs of spare parts and labour [162]; as well as the Norwegian company Norske Skog which relies on online monitoring to minimise costly unplanned replacements of pump bearings. Extensive vibration monitoring has been carried out in paper mills for many years, but there is still a considerable scope for improvement both in technique and management of CM [107]. The use of vibration monitoring generally reduces the number of failures, but it may also, at the same time, over-maintain the plant in question. Appropriate forecasting is required here to support the condition-based decision making.

Irving Pulp and Paper produces 325 000 tons of wood pulp each year at its facilities in Saint John, New Brunswick, Canada, employing approximately 425 people. The pulp produced is then used for making paper, tissue and corrugating medium. The stress on pulp and paper market prices, combined with continuously increasing production costs, is creating a situation in which the mill must maximise productivity and cost effectiveness to keep margins from eroding. Therefore, reducing equipment failures and maximising machinery uptime are among the key objectives of the mill.

The case study started with a review of the mill’s available data. Centrifugal pumps are used extensively for pumping the various liquids used in the paper making process from one processing station to another. This study looks at the Gould 3175L centrifugal pumps which had a high incidence of unpredictable failures. These pumps are critical equipments that operate 24/7 except during planned maintenance shutdowns, which last for two to three days each time and happen twice a year. These pumps were originally deliberately oversized for the work they are required to do in order to anticipate possible capacity expansion. To compensate for the sub-optimal workload, the pumps were run below their best efficiency point by throttling the discharge flow. This practice caused excessive load on the pump bearings and greatly contributed to the high failure rate of bearings. Pump failures can incur costly downtime. However, the pump workload depends on the amount of pulp required by the market. Therefore, instead of changing the workload, Irving
Pulp and Paper has modified the pumps on-site to reduce the stress on the bearings, and is using CM to reduce unplanned stoppages.

6.2 The Data

The pumps operate with the same constant load and speed. Vibration readings are collected using a hand-held device placed at eight locations on the pump. Of the eight measurement points, four are horizontal (perpendicular to pump rotation), three vertical (perpendicular to pump rotation), and the remaining one axial (along the same line as rotation axis). The hand-held device stores vibration readings for subsequent uploading into a master database. The vibration signal is pre-processed into five frequency bands, an overall summary of the five bands, and an acceleration value. These seven features are reported at each of the eight measurement points, resulting in a total of 56 features for each measurement.

For this case study, 32 histories (obtained between early 2001 to early 2005) were available. Of the 32 histories, 12 ended in rolling element bearing failure and six in seal failure. The remaining 14 histories were suspended since the pumps were still operating normally when the data were obtained. As the failure mode to be considered in this study was bearing failure, the 6 mechanical seal failure histories functioned as suspended histories. The seal failures did not affect the vibration readings and were found to be completely random. Covariate analysis using the software EXAKT [163] confirmed that none of the 56 features was significantly related to the seal failures. Rolling element bearing problem is the main concern for pulp and paper mills [107, 162, 164, 165]. They are important components in paper mills, not only because of the large quantity of them installed in paper machines, but also due to their role in relation to the quality of the paper produced [107].

All the bearings in this case study were identical as they were of the same dimensions, made of the same material, and worked under the same operating conditions. Using the EXAKT covariate analysis, two features were found to be
significantly related to bearing degradation. These two features, namely P1H_Par5 and P1V_Par5, corresponded to the 5th frequency band of the horizontal and vertical measurements at the far end of the pump from impeller. Indeed, the problematic bearings were each located at that far end of the pump. Irving vibration specialists also confirmed that these two features were sensible variables to be considered. However, being located at the same bearing end and corresponding to the same frequency band, the two features were highly correlated and might introduce redundancy in the description of the bearing degradation. To reduce the feature dimension and, in turn, reduce the ANN training time, only P1V_Par5 was used since it was found to be the most significant covariate.

Figure 6-1 shows the plot of P1V_Par5 of the 12 failed bearings. The bearing lives varied from 63 days to 1468 days, showing a typical stochastic nature of survival distribution. The plot also reveals an increasing trend in the condition indicator with bearing age in all cases. Although the degradation path in each case is different, the overall degradation pattern is quite similar. The condition index remained quite flat throughout the earlier stage of the bearing life, and then increased at various rates during the failing stage.
Since the pumps were not always run to an actual breakdown and failing components were sometimes opportunistically replaced after they were found faulty, not all the “failure” histories have reached complete failure. They might have still been able to operate for a period of time before a breakdown could happen. However, information about the failure condition and defect severity at the time of replacement was not recorded. The difficulty here lies in identifying the histories that have reached complete failure and histories that end in informative suspension. After examining the feature trending plots, Histories No.14 and No.24 were found to terminate at noticeably lower feature value than the others. Therefore, these two histories are taken as informative suspension cases. The health index of History

---

**Figure 6-1: Vibration histories of the 12 failed bearings**
No. 14 was determined as 0.3 and that of History No. 24 was assigned a value of 0.2. Thus, there are only seven complete failure histories remaining.

### 6.3 Prognostic Modelling

Three of the seven failure histories were reserved as test-sets and the remaining histories (5 failures, 2 informative suspensions, 14 suspensions) were assigned for modelling and network training (Table 6-1). Note that these real life data indeed contain even more suspended histories than the data simulated in the previous chapter, emphasising the importance of the correct modelling of suspended histories. Also, since the pumps operate 24/7 with constant speed, there was no need to convert the calendar age to number of operating cycles and, thus, the number of days was directly used as the operating age.
Table 6-1: Pump histories used for model training and testing

<table>
<thead>
<tr>
<th>History No.</th>
<th>Time of failure/suspension (day)</th>
<th>21</th>
<th>30</th>
<th>Training datasets that end in failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>635</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>650</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>742</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>723</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>504</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>405</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>506</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>960</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>664</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>693</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>440</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1402</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>992</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1398</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1468</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1427</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>446</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>473</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>283</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>601</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>1246</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>964</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The survival function for the entire training histories estimated using the standard KM estimator is depicted in Figure 6-2.
The data collection intervals were not equally spaced. The measurement frequency ranged from around 25 to 60 days during normal operation. As a bearing was failing, however, its condition would usually be taken more often, sometimes as frequently as once every several days. As mentioned, one solution is to obtain the continuous function by constructing a regression line over the points of condition measurements. However, less computationally intensive solutions are available. For this test, condition values were linearly interpolated so that the measurement points were equally spaced at 10 days. Three example histories are shown in Figure 6-3 and Figure 6-4. The interpolated curves were almost identical to the original curves.

Figure 6-2: The standard KM estimation of survival probability for the training histories
Figure 6-3: Original and interpolated condition data of History No.10
Failure History No.11 was too short to be included in ANN training, so was only used as event data for estimating survival probability. There were then only six failure histories left for model training.

The model performance evaluation as described in the previous chapter was carried out, this time using these real life data. The performance of the proposed model was compared to the performance of the two FFNN control models that neglected suspended data, of an RNN that performed single-step-ahead time series prediction,
and of the traditional Weibull model that was based solely on event data. The failure threshold for the RNN was set to be 0.6. The test consisted of 3 assessments (Table 6-2). In Assessment I, all 6 complete histories and 16 suspended ones were made available for model training. In Assessment II, only 3 complete histories and the 16 suspended ones were used for training. In the last assessment, only 1 complete failure history and 16 suspended histories were available for training.

Table 6-2: Histories used for training in Assessments I, II and III

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Number of Histories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Failure</td>
</tr>
<tr>
<td>I</td>
<td>6</td>
</tr>
<tr>
<td>II</td>
<td>3</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
</tr>
</tbody>
</table>

The practical considerations when choosing the ANN structures (as described in Subsection 5.2.3 in the previous chapter) were taken. The FFNNs used for this real life data analysis each had 11 input nodes (consisting of the “current” asset condition value and 10 delayed values) and 5 output nodes (predicting 5 intervals ahead). They each possessed 1 single layer of 15 hidden nodes. The RNN had 9 hidden nodes and was capable of predicting 1 step ahead. The network training and transfer functions remain the same as those used in the simulation test.

6.4 Assessment Results

The three sets of prediction results from each of the five prognostic models were analysed individually. As with the simulated data analysis, the predicted failure time was identified by noting the first output unit that estimated a survival
probability of less than 0.5. Subsections 6.4.1 - 6.4.5 discuss the prognosis results of each prognostic model respectively. As there are five models to be discussed and each model has nine sets of prediction results (three test-sets × three assessments), only the results of the first test-set in Assessment I will be discussed in detail. The other results will be summarised and compared in Subsection 6.4.6.

6.4.1 Results of Assessment I by the Proposed Model

It was observed that the survival probability estimated by the proposed model closely matched the actual failures. Figure 6-5 shows the first test-set – History No.30 – in which the actual failure time was at $t=600$; and
Table 6-3 shows the corresponding prediction results. In this test, the \( n \)th row of prediction output at each time step represents the probability that the unit will survive the \( n \)th next time step. The predicted survival probabilities closely matched the actual degradation trend.

![Degradation index of pump bearing in the 1st test-set (History No.30)](image)

**Figure 6-5:** Degradation index of pump bearing in the 1st test-set (History No.30)
Figure 6-6 shows the interpolated input data and the graphical representation of predicted survival probability at selected time steps. The survival probability was high (above 0.8) and had a stable trend during earlier lifetime of the bearing (from the start of operation to approximately 400 days, as shown in Table 6-3 and Box A in Figure 6-6). The survival probability began to drop slightly more at around 430 days (Box B in Figure 6-6), suggesting the initiation of a defect. Another interesting observation which can be seen in Figure 6-6 is that, despite the sudden stop of the climb in the vibration RMS value at around $t = 500$ days, the survival probability was forecasted to drop at an increasing rate (Box C in Table 6-3: Survival probabilities predicted by the proposed model for the 1st test-set (History No.30) in Assessment I

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>Surv. probability in 2nd following interval $\hat{S}_{t+2\Delta t}$</th>
<th>Surv. probability in 3rd following interval $\hat{S}_{t+3\Delta t}$</th>
<th>Surv. probability in 4th following interval $\hat{S}_{t+4\Delta t}$</th>
<th>Surv. probability in 5th following interval $\hat{S}_{t+5\Delta t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t=110$</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>$t=120$</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>$t=130$</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>$t=140$</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>$t=150$</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>$t=160$</td>
<td>0.84</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>$t=170$</td>
<td>0.84</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>$t=180$</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>$t=190$</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>Failed at $t=600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6-6). This observation suggests that the prognostic model may have learned to capture the non-linear relationship between the condition index and the actual health state of the monitored bearing. This capability makes such models much more robust than models that directly use condition index to represent asset health (e.g. time series prediction models). At $t = 560$ days, the survival probability was predicted to fall from 0.54 on day 570, to 0.09 on day 610 (Box D in Figure 6-6). Towards the end of the bearing life, at $t = 580$ days (Box E in Figure 6-6), the survival probability was forecasted to fail slightly below 0.5 in the following 10 days and to below 0.3 in 30 days from time of prediction. Even though the vibration RMS started falling after $t = 560$ days, the survival probability forecasted for the immediate next 10 days at $t = 570$, 580 and 590 remained below 0.5. This again confirmed that the model was able to learn the failure pattern, instead of simply modelling the survival probability to be inversely proportional to the condition index.
It is noted, however, that if a survival probability of 0.5 was used as the failure threshold (as was determined for this study), the model underestimated the failure time. The first output with a value below 0.5 was produced at $t = 530$ in the 5th row ("0.42", highlighted in Figure 6-6: Graphical representation of the prediction output by the proposed model at selected time steps for the 1st test-set (History No.30) in Assessment I).
Table 6-3), which means the bearing was forecasted to fail in the 5\textsuperscript{th} next interval, i.e. $t = 580$ days. However, the failure did not occur till $t = 600$ days. The error is considered small in relation to the whole lifetime of the bearing:

Prediction error
\[
= \frac{\text{Actual failure time} - \text{Predicted failure time}}{\text{Item lifetime}} \\
= \frac{600 - 580}{600} \\
= 0.033 \text{ (or 3.3\%)}
\]

This underestimation, however, might be due to the supposed “complete failure” training datasets still having a certain amount of remaining useful life at replacement. Even though the failed bearings were confirmed by the maintenance engineer to be noticeably damaged at replacement, there was no information as to how much longer the bearings could have run before an actual breakdown would occur. This short period of time discrepancy may have created a slight bias in the failure data modelling. The bearing in the test-set in question might have been run to a higher level of defect severity before being replaced and, therefore, the failure point seemed to be slightly later in the lifetime than the normal failure point that the proposed ANN had learned to recognise. In fact, History No.30 indeed has a longer period of decreasing vibration RMS value at the end of the bearing life, compared to the training datasets. This observation may suggest that the bearing in History No.30 might indeed have been left to run to a higher stage of damage than the bearings in the failure training datasets.

6.4.2 Results of Assessment I by Model A

As expected, model A (the FFNN with the same ANN structure and training method as the proposed model but which modelled suspensions as failures), underestimated the failure times even more than the proposed model. Table 6-4 shows the prediction results for the first test-set – History No.30 – and Figure
Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data

6-7 shows the interpolated input data and the graphical representation of predicted survival probability at selected time steps.

Table 6-4: Predicted survival probabilities produced by Model A for the 1st test-set (History No.30) in Assessment I

<table>
<thead>
<tr>
<th>t=110</th>
<th>t=120</th>
<th>t=130</th>
<th>t=140</th>
<th>t=150</th>
<th>t=160</th>
<th>t=170</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surv. probability in 1&lt;sup&gt;st&lt;/sup&gt; following interval $\hat{S}(t+\Delta t)$</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.73</td>
<td>0.72</td>
</tr>
<tr>
<td>... 2&lt;sup&gt;nd&lt;/sup&gt; following interval $\hat{S}(t+2\Delta t)$</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>... 3&lt;sup&gt;rd&lt;/sup&gt; following interval $\hat{S}(t+3\Delta t)$</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.70</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td>... 4&lt;sup&gt;th&lt;/sup&gt; following interval $\hat{S}(t+4\Delta t)$</td>
<td>0.67</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>... 5&lt;sup&gt;th&lt;/sup&gt; following interval $\hat{S}(t+5\Delta t)$</td>
<td>0.66</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=180</th>
<th>t=190</th>
<th>t=200</th>
<th>t=210</th>
<th>t=220</th>
<th>t=230</th>
<th>t=240</th>
<th>t=250</th>
<th>t=260</th>
<th>t=270</th>
<th>t=280</th>
<th>t=290</th>
<th>t=300</th>
<th>t=310</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.72</td>
<td>0.72</td>
<td>0.71</td>
<td>0.70</td>
</tr>
<tr>
<td>0.69</td>
<td>0.69</td>
<td>0.70</td>
<td>0.70</td>
<td>0.71</td>
<td>0.71</td>
<td>0.72</td>
<td>0.71</td>
<td>0.70</td>
<td>0.70</td>
<td>0.69</td>
<td>0.68</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.69</td>
<td>0.69</td>
<td>0.68</td>
<td>0.68</td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.68</td>
<td>0.68</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td>0.65</td>
<td>0.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=320</th>
<th>t=330</th>
<th>t=340</th>
<th>t=350</th>
<th>t=360</th>
<th>t=370</th>
<th>t=380</th>
<th>t=390</th>
<th>t=400</th>
<th>t=410</th>
<th>t=420</th>
<th>t=430</th>
<th>t=440</th>
<th>t=450</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.69</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.66</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>0.70</td>
<td>0.70</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.69</td>
<td>0.68</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.65</td>
<td>0.64</td>
<td>0.63</td>
</tr>
<tr>
<td>0.69</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.65</td>
<td>0.64</td>
<td>0.62</td>
</tr>
<tr>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.64</td>
<td>0.64</td>
<td>0.63</td>
<td>0.62</td>
<td>0.61</td>
</tr>
<tr>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.65</td>
<td>0.65</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.63</td>
<td>0.62</td>
<td>0.61</td>
<td>0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=460</th>
<th>t=470</th>
<th>t=480</th>
<th>t=490</th>
<th>t=500</th>
<th>t=510</th>
<th>t=520</th>
<th>t=530</th>
<th>t=540</th>
<th>t=550</th>
<th>t=560</th>
<th>t=570</th>
<th>t=580</th>
<th>t=590</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.64</td>
<td>0.62</td>
<td>0.59</td>
<td>0.57</td>
<td>0.55</td>
<td>0.53</td>
<td>0.52</td>
<td>0.52</td>
<td>0.49</td>
<td>0.48</td>
<td>0.46</td>
<td>0.45</td>
<td>0.43</td>
<td>0.46</td>
</tr>
<tr>
<td>0.60</td>
<td>0.56</td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.51</td>
<td>0.48</td>
<td>0.42</td>
<td>0.37</td>
<td>0.36</td>
<td>0.37</td>
<td>0.38</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td>0.59</td>
<td>0.56</td>
<td>0.52</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
<td>0.48</td>
<td>0.44</td>
<td>0.39</td>
<td>0.35</td>
<td>0.30</td>
<td>0.25</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>0.57</td>
<td>0.54</td>
<td>0.51</td>
<td>0.50</td>
<td>0.47</td>
<td>0.46</td>
<td>0.42</td>
<td>0.36</td>
<td>0.29</td>
<td>0.26</td>
<td>0.23</td>
<td>0.21</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>0.55</td>
<td>0.51</td>
<td>0.50</td>
<td>0.45</td>
<td>0.44</td>
<td>0.43</td>
<td>0.40</td>
<td>0.33</td>
<td>0.25</td>
<td>0.17</td>
<td>0.11</td>
<td>0.11</td>
<td>0.20</td>
<td>0.25</td>
</tr>
</tbody>
</table>

failed at t=600
The predicted survival probability was lower than that estimated by the proposed model since the start of the bearing operation. The probability value falls below 0.5 as early as around \( t = 490 \) days in the 5\(^{th}\) row, meaning that the monitored bearing was expected to fail within the following 50 days. However, the actual failure only occurred 110 days later. The prediction error produced by this model is twice as large as that produced by the proposed model.

Prediction error

\[
\text{Prediction error} = \frac{\text{Actual failure time} - \text{Predicted failure time}}{\text{Item lifetime}}
\]

\[
= \frac{600 - 540}{600}
\]

\[
= 0.1 \text{ (or 10%)}
\]
It was noticed that Model A has also appeared to learn the non-linear relationship between the given vibration data and the actual item health. This capability can be observed at, for example, around day 500. The model was able to recognise that the bearing health was still deteriorating, even though the increment rate of the vibration value dropped during that period. This observation was not surprising since Model A had the same network architecture as the proposed model. The only difference is that this model was trained based on the false assumption that all bearings have reached failure at replacement, regardless of the damage severity. This weakness could be observed not just in the underestimation of failure time, but also in the fluctuations of the predicted survival probability close to failure. The false training might have clouded the signs of an impending failure veiled in the condition data – signs that should have been picked up by the network.

6.4.3 Results of Assessment I by Model B

The prognosis results of Model B (the FFNN with the same structure and training function as the proposed model but omitted suspended histories in training) for each test-set were also examined. Table 6-5 shows the prediction results for the first test-set and Figure 6-8 shows the interpolated input vibration data, along with the predicted survival graph at selected time steps.
Table 6-5: Predicted survival probabilities produced by Model B for the 1st test-set (History No.30) in Assessment I

| Surv. probability in 1\textsuperscript{st} following interval $\tilde{S}_{(t+\Delta t)}$ | t=110 | t=120 | t=130 | t=140 | t=150 | t=160 | t=170 | t=180 | t=190 | t=200 | t=210 | t=220 | t=230 | t=240 | t=250 | t=260 | t=270 | t=280 | t=290 | t=300 | t=310 |
| 0.89 | 0.89 | 0.89 | 0.89 | 0.90 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 |
| 0.88 | 0.88 | 0.87 | 0.88 | 0.89 | 0.91 | 0.91 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 |
| 0.87 | 0.86 | 0.86 | 0.87 | 0.88 | 0.88 | 0.89 | 0.89 | 0.90 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 |
| 0.86 | 0.85 | 0.85 | 0.86 | 0.87 | 0.88 | 0.89 | 0.89 | 0.90 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 |
| 0.85 | 0.85 | 0.84 | 0.85 | 0.86 | 0.87 | 0.88 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 |

<table>
<thead>
<tr>
<th>t=320</th>
<th>t=330</th>
<th>t=340</th>
<th>t=350</th>
<th>t=360</th>
<th>t=370</th>
<th>t=380</th>
<th>t=390</th>
<th>t=400</th>
<th>t=410</th>
<th>t=420</th>
<th>t=430</th>
<th>t=440</th>
<th>t=450</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
<td>0.90</td>
<td>0.91</td>
<td>0.90</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>0.84</td>
<td>0.85</td>
<td>0.87</td>
<td>0.89</td>
<td>0.92</td>
<td>0.91</td>
<td>0.90</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.85</td>
<td>0.88</td>
<td>0.90</td>
<td>0.93</td>
<td>0.91</td>
<td>0.89</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>0.82</td>
<td>0.83</td>
<td>0.84</td>
<td>0.86</td>
<td>0.88</td>
<td>0.91</td>
<td>0.90</td>
<td>0.89</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t=460</th>
<th>t=470</th>
<th>t=480</th>
<th>t=490</th>
<th>t=500</th>
<th>t=510</th>
<th>t=520</th>
<th>t=530</th>
<th>t=540</th>
<th>t=550</th>
<th>t=560</th>
<th>t=570</th>
<th>t=580</th>
<th>t=590</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>0.82</td>
<td>0.80</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.78</td>
<td>0.76</td>
<td>0.69</td>
<td>0.64</td>
<td>0.59</td>
<td>0.57</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>0.74</td>
<td>0.70</td>
<td>0.67</td>
<td>0.66</td>
<td>0.68</td>
<td>0.71</td>
<td>0.74</td>
<td>0.71</td>
<td>0.61</td>
<td>0.56</td>
<td>0.52</td>
<td>0.49</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td>0.70</td>
<td>0.65</td>
<td>0.61</td>
<td>0.59</td>
<td>0.62</td>
<td>0.65</td>
<td>0.69</td>
<td>0.66</td>
<td>0.54</td>
<td>0.43</td>
<td>0.29</td>
<td>0.21</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>0.66</td>
<td>0.61</td>
<td>0.56</td>
<td>0.54</td>
<td>0.57</td>
<td>0.62</td>
<td>0.67</td>
<td>0.65</td>
<td>0.49</td>
<td>0.33</td>
<td>0.15</td>
<td>0.08</td>
<td>0.08</td>
<td>0.22</td>
</tr>
<tr>
<td>0.63</td>
<td>0.56</td>
<td>0.51</td>
<td>0.50</td>
<td>0.53</td>
<td>0.58</td>
<td>0.62</td>
<td>0.49</td>
<td>0.34</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
</tr>
</tbody>
</table>

failed at t=600
The failure time estimated by this model was closer to the actual failure time compared to the other models. The predicted survival probability value first dropped to below 0.5 in column $t = 530$ in the 5th row, meaning that the monitored bearing was forecasted to fail around day 580, which was 10 days before the actual failure. The prediction error is quite small in comparison to the previous two models.

Prediction error

$$= \frac{\text{Actual failure time} - \text{Predicted failure time}}{\text{Item lifetime}}$$

$$= \frac{600 - 580}{600}$$

$$= 0.033 \text{ (or 3.3\%)}$$

Figure 6-8: Graphical representation of the prediction output by Model B at selected time steps for the 1st test-set (History No.30) in Assessment I
The reasonable prediction accuracy might be due to the relatively large number of failure histories available for training in this assessment. If the six failure histories were sufficient for training the network to recognise failure patterns, the network would be able to predict well although it did not utilise the suspended histories as training datasets. Suspended histories can not be modelled as accurately as the failure histories since the actual survival status past the point of suspension is not known. However, the prediction accuracy of Model C might drop when there are fewer failure histories available for training.

Although the failure time predicted by Model B was close to the actual failure time, the estimated survival probability was never monotonously decreasing. This deficiency greatly reduced the reliability of the model. Indeed, the model appeared to be inconsistent as the prediction results for the other two test-sets were not as accurate.

6.4.4 Results of Assessment I by Model C

Figure 6-9 presents the prediction results of the one-step-ahead time series RNN model, which commonly uses the condition index to represent the item health in prognostic applications. At the first several time steps, the predicted value was far from the actual value as the network had not accumulated enough delayed values to start performing accurate prediction. After about five time steps (50 days) the network had adapted itself and started to produce closely matched value. However, it could be observed that the network was only trying to adapt to the previous actual value and, therefore, the predicted value lagged behind the actual value by one time step most of the time.
6.4.5 Results of Assessment I by Model D

The traditional Weibull model estimated a MTTF of 1614.7 days for the train datasets. For this simulation test, the maximum likelihood estimates of the shape parameter, $\beta$, and the scale parameter, $\eta$, were 1767.90 and 1.38. The corresponding Weibull plot is depicted in Figure 6-10.

![Figure 6-9: The RNN prediction results for the 1st test-set (History No.30) in Assessment I](image-url)

- 175 -
The Weibull function had a positively skewed distribution since the data were highly censored. The estimated MTTF was far longer than the actual failures in the three test-sets (in which failure occurred on day 601, day 1246 and day 964 respectively). The MLE of the shape and scale parameters seemed problematic when estimating from small and highly censored samples. Even though the MLE approach has long been regarded as a satisfactory reference technique, it can be imprecise and sometimes unreliable for data that are limited or highly censored [166]. Bayesian inference has been proposed for estimating industrial reliability data, since these data are usually limited and highly censored, and industrial experts may have prior information on the distribution parameters [167]. However, direct Bayesian analysis for Weibull distribution is complex – especially the estimation of distribution shape parameter.

---

**Figure 6-10: Weibull plot by Model D**

The Weibull function had a positively skewed distribution since the data were highly censored. The estimated MTTF was far longer than the actual failures in the three test-sets (in which failure occurred on day 601, day 1246 and day 964 respectively). The MLE of the shape and scale parameters seemed problematic when estimating from small and highly censored samples. Even though the MLE approach has long been regarded as a satisfactory reference technique, it can be imprecise and sometimes unreliable for data that are limited or highly censored [166]. Bayesian inference has been proposed for estimating industrial reliability data, since these data are usually limited and highly censored, and industrial experts may have prior information on the distribution parameters [167]. However, direct Bayesian analysis for Weibull distribution is complex – especially the estimation of distribution shape parameter.
6.4.6 Overall Model Comparison for All Assessments

The penalty points of the proposed model and Models A, B, C and D are presented in Table 6-6.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Proposed</th>
<th>A (models suspensions as failures)</th>
<th>B (excludes suspensions from training)</th>
<th>C (one-step-ahead time series prediction)</th>
<th>D (traditional Weibull model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (6F,16S)</td>
<td>0.541</td>
<td>1.001</td>
<td>1.601</td>
<td>1.119</td>
<td>10.200</td>
</tr>
<tr>
<td>II (3F,16S)</td>
<td>0.601</td>
<td>1.935</td>
<td>1.635</td>
<td>1.119</td>
<td>10.350</td>
</tr>
<tr>
<td>III(1F, 16S)</td>
<td>0.821</td>
<td>1.768</td>
<td>6.385</td>
<td>1.152</td>
<td>25.950</td>
</tr>
</tbody>
</table>

The performance results reinforce the findings reported in the previous chapter. The proposed model consistently received the lowest penalty points. Model A received considerably higher penalty points because it underestimated the time to failures. Model B did not perform well, especially in the last assessment where only one failure history was used for training. The predicted degradation produced by Model C often lagged behind the actual process and, therefore, failure usually occurred before the predicted time. This undesired lag and short prediction horizon contributed to the penalty for the model. It was also found that the variation in the number of training datasets did not have effects on the performance of this time series prediction network. This might be because it basically performs single-step-ahead adaptation. Model D was the weakest model in this study as the MTTF that it estimated was far from the actual lifetime of the test-set bearings in each assessment. This model was based solely on reliability data, which contain no information on the bearing degradation process. As a result, the model was severely penalised – even though it facilitated long prediction horizon.
6.5 Summary

In this chapter, the proposed prognostic model was successfully tested through an industry case study. The condition and event data of centrifugal pumps at Irving Pulp and Paper Mill were used to test and compare the proposed model with four other control models. It was shown that a non-parametric, data-driven approach that integrates both population reliability characteristics and CM information is able to simultaneously reduce the impact of model errors and provide a balanced performance in terms of prediction horizon and accuracy. Other existing prognostic models that also integrate reliability information into asset condition prediction are mostly parametric models. The most widely known model is the PHM model discussed in Chapters 3 and 4. However, PHM models the instantaneous failure rate of a unit solely as a function of time and the instantaneous value of asset condition indices. The failure rate of a unit is assumed to be independent of the past condition values. A justification for ignoring this possible dependence is that it makes the model intolerably complex [168]. It was stated that, even without considering the past condition values, the model was already complex enough due to the estimation of various model parameters. The proposed prognostic model, on the other hand, is able to consider a series of asset condition values, while remaining computationally inexpensive due to the ANN parallel computation capability.

Even though (in this study) the proposed model performed better than similar models that do not model suspended CM histories and a commonly used time series prediction network as well as a traditional Weibull model, it has not reached its highest potential accuracy. The pump CM data available at the paper mill only consisted of several pre-processed vibration features collected at varying intervals. The prediction accuracy might be improved by including more consistent trending data, along with other condition information such as oil particle count. The proposed model is suitable for on-line implementation, since the neural network’s noise-insensitivity and high-parallelism enable reasonably fast and accurate prediction – even in the presence of uncertain data and measurement errors. The neural computation paradigm, if it is correctly utilised in prognostics and does not
have its use limited to the time series prediction approach alone, can be an extremely powerful tool to analyse complex datasets.

It is acknowledged, however, that there may never be sufficient data to completely characterise the true PDE of asset condition empirically and to train the ANN with all possible failure pattern examples. Consequently, prognostic models that employ PDE and ANN may be flawed without an opportunity for validation. Nonetheless, data capture and fusion techniques are becoming increasingly advanced. The proposed approach, combined with sufficient high quality data, represents a compelling concept for longer-range data-driven prognosis, and for utilising available asset condition information more fully and accurately.
CHAPTER 7  CONCLUSION AND FUTURE WORK

This chapter first summarises the contributions of this work as the answers to the research questions raised. Suggestions for future work are outlined in the second section. The chapter concludes with some final remarks about the implications of the research findings.

7.1 Summary

The ability to forecast machinery failure is vital to reducing maintenance costs, operation downtime and safety hazards. Recent advances in CM technologies have given rise to a number of prognostic models that attempt to forecast machinery health based on condition data. This thesis reported a comprehensive and critical review of the existing machinery health prognostic techniques, in addition to an overview of CBM technologies and their evolution. The extensive literature review has shown that, although the existing models have advanced the development of machinery prognostics, the following aspects need to be further investigated before prognostic systems can be reliably applied in real-life situations:
1. The existing prognostic models have not directly modelled and fully utilised suspended CM data, which are very common in practice, since organisations would rarely let their assets run to failure. Treating historical suspensions as failures in prognostic modelling would result in underestimation of remaining asset life. On the other hand, excluding suspended CM histories from modelling would worsen the problem of data unavailability.

2. The existing models have not effectively integrated population characteristics and CM information for longer-range prognosis. While CM data reflect the state of individual operating units, they do not replace reliability data that reflect population characteristics.

3. In the existing prognostic models, CM data are taken to indicate the health of a monitored unit. However, the measured condition indices do not always deterministically represent the actual health of the monitored asset. There is not yet a model that deduces the non-linear relationship between an asset’s actual survival status and the acquired CM data.

4. The existing models rely on a large number of assumptions, such as assumptions about the underlying system physics, failure distribution or degradation pattern. In practice, formulating an accurate parametric model is very difficult due to poorly established underlying physics, difficulty in identifying defect type or geometry, complex interactions between the monitored component and other components or the operating environment, as well as the stochastic nature of failure and degradation patterns.

A novel prognostic model has been proposed in this work to address the above challenges. Instantaneous reliability of historical items is first calculated using a variation of KM estimator and a degradation-based failure PDE method. The estimated reliability is used as the training targets for an FFNN. The trained FFNN will be capable of estimating the future survival probability of the monitored item, given the corresponding CM indices. This work presents an approach that:
1. Includes suspended CM histories in model training examples. Suspended data are fully utilised and amalgamated directly into prognostic modelling through the adapted KM estimator, not by using an artificial cut-off time, but rather by using the information that the item has survived up to the last observed time and the probability of the population surviving the following intervals. The inclusion of suspended CM data in prognostic modelling avoids underestimation of remaining asset life and mitigates the data scarcity problem.

2. Integrates all reliability data (including event data that do not come with corresponding CM data) into prognostic modelling through a KM estimator; and extracts population characteristics and conditional reliability from CM histories instead of from reliability event data. These procedures not only effectively integrate population reliability and unit CM information for longer-range prediction, but also enable reliability to be calculated based on the true degradation processes of historical assets, instead of on the historical time-to-failure data alone.

3. Uses an ANN to recognise the non-linear relationship between actual asset health and measured condition data. Like human experts, the proposed ANN learns from experience. However, the use of ANNs enabled permanent and transferable storing of knowledge. ANNs have also been proven to outperform traditional statistical techniques in non-linear data fitting, working with noisy data and fast processing due to their high parallelism. The model was able to learn to recognise how unit degradation is veiled in the non-deterministic changes in CM measurements, and to disregard fluctuations caused by non-deterioration factors.

4. Uses a non-parametric approach to performing prognosis based on the true data obtained. The ability to predict without relying on assumptions such as those concerning system physics properties, degradation pattern, failure distribution and failure threshold avoids the potentially large errors brought about by making incorrect assumptions about those properties.
The proposed model, therefore, meets the objectives of addressing the challenges identified in the literature review. Computer-simulated data and industry pump vibration data were used for model training, testing and comparison. The results verified that the proposed model performs better than models that do not include suspended histories and/or population characteristics in prognostic modelling. The difference in performance was most significant when complete failure data were limited.

The proposed model has the potential of being applied to other types of assets with different degradation patterns. Example applications include the forecast of tool wear, structural corrosion and electronic component failure. The proposed model is data-driven and requires no system physics knowledge or degradation pattern assumptions.

7.2 Future Work

Since the model validation has been undertaken in collaboration with industry, several new research issues have been identified and are listed as follows:

1. Although the centrifugal pumps data used for this study were not affected by maintenance, this is not always the case in many other applications. Maintenance actions such as change of lubricant can disrupt the progressive degradation trend. However, none of the existing prognostic techniques found in the literature have modelled these effects. In the field of classical reliability, however, the survival probability of repairable systems has been extensively modelled. A comprehensive review of existing reliability models and maintenance policies can be found in Ref. [176]. These techniques for modelling the effects of maintenance actions on system reliability may be enhanced with CM data that actually reflect the system’s online health state.
2. The centrifugal pumps considered in this work had a constant operating condition. In many situations, however, machines are subject to varying operating conditions. This form of variation is a major contribution to the change in the energy of measured CM signals. Therefore, it is necessary to extend this prognostics research effort to ensure the proposed model is only sensitive to the changes in condition measurements caused by asset deterioration, and insensitive to the influence of changes in operating condition. In the literature on fault detection, load-independent features have been extracted from CM data. These load-independent features need to be validated for application in prognostics.

3. The proposed model should be applied to other types of assets to evaluate the possibility of generalising the approach to forecasting a variety of failure patterns. One of the most obvious merits of the proposed model is that it is completely data-driven and requires no system physics knowledge or degradation pattern assumptions. Possible applications include the prediction of tool wear, structure corrosion and electronics failure.

4. A decision model can be developed to relate the reliability estimated by the proposed prognostic model to a maintenance decision policy. The optimal decision of whether to replace a unit will be made by minimising a cost function that consists of the preventive replacement cost, failure replacement cost and optimal risk threshold.

5. The model validation results have demonstrated that the proposed model can be beneficial to industry. However, application of this model might appear complex and overwhelming to maintenance personnel who often do not have in-depth knowledge about the algorithms. Appropriate software tools can be developed to enable convenient implementation of the model.
7.3 Final Remarks

In today’s market, the ability to remain competitive is the difference between success and failure. Organisations strive to survive in the ever increasing cost pressure and competitiveness spiral that demands continually increased effectiveness and returns on investment. In many industries, equipment is the largest investment an organisation makes. Most of the time, unfortunately, the potential of this equipment is not fully realised.

The majority of organisations today are focusing on maintenance and reliability as costs rather than as competitive weapons. As mentioned in the introductory chapter of this thesis, maintenance costs have been increasing exponentially to a level that organisations can no longer absorb. Many organisations have moved their operations to other countries for lower operating costs. Is it the best solution to shift operations offshore instead of properly utilising the hard assets in which the organisations have already invested? Perhaps not. As stated at the start of this chapter, maintenance may be the most overlooked opportunity within industry today. Organisations and nations can regain competitiveness through improved asset utilisation, and this begins with the ability to forecast asset condition and optimise maintenance.

This work has demonstrated that machinery health can indeed be forecasted. The proposed prognostic technique – together with ongoing advances in sensors and data-fusion techniques, and increasingly comprehensive databases of asset condition data – holds the promise of increased asset availability, maintenance cost effectiveness, operational safety and, ultimately, organisation competitiveness.
REFERENCES


Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data


Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data


Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data


Intelligent Prognostics of Machinery Health Utilising Condition Monitoring Data


