Bayesian Approaches for Complex System Prognostics

The University Of Sheffield.

Martha Arbayani Bin Zaidan

Department of Automatic Control and Systems Engineering

University of Sheffield

A thesis submitted for the degree of

Doctor of Philosophy

March 2014

I would like to dedicate this thesis to

my parents, ZAIDAN & RAHMAH

and

my family, SUCI & NACITHA

Acknowledgements

All praise and glory is due to Almighty Allah, the sole Lord of the universe, whose mercy and blessings have been bestowed constantly upon me. Peace and blessings be upon His final messenger Muhammad.

I would like to express my deep felt gratitude to my supervisors, Professor Robert F. Harrison and Professor Peter J. Fleming for giving me the opportunity to pursue my Doctoral research under their guidance. I specially thank Professor Robert F. Harrison for his helpful insights on Bayesian methods as well as machine learning and Professor Peter J. Fleming for providing constructive criticism about my work. I appreciate and really value all their feedback and guidance.

I want to express my thanks to my thesis examiners: Professor Christopher Bailey and Dr. Osman Tokhi for their positive comments and corrections in my thesis.

I am grateful to Rolls-Royce plc and Engineering & Physical Sciences Research Council (EPSRC) for providing me the necessary financial support through Dorothy Hodgkin Postgraduate Award (DHPA). Interaction with Rolls-Royce representatives was really helpful in shaping the outcome of this research.

I would also like to thank the members of Rolls-Royce UTC for control and monitoring systems who gave so much of their time to help me, Rishi Relan for the lengthy discussions on machine learning and information theory, Andy Mills for discussion in health monitoring technologies and being able to solve just about anything administrative. In addition, thanks are due to Vignesh Venugopalan, Rui Wang, Hasanain Bashar and friends who made my stay in Sheffield so enjoyable. The time spent with them will leave with me several fond memories of Sheffield, and also to my friends as well as colleagues from Indonesia who have supported me along the way.

I got many helps during my thesis writing from colleagues at the University of Sheffield. I wish to thank Ubaid Bavakhan for assisting me extensively in LaTeX, Ali Zughrat for

helping in proofreading of my thesis, Mehmet Ali Eroglu for recommending me working in Mathematics and Statistics Help (MASH), Muneer Omar for providing me place to write the thesis and Abdallah Hussin for providing me the place to stay at the end of my PhD period. I extend my thanks to my faithful friends and neighbours, who are indeed my second family in UK.

I am indebted to my family; mother and father. Without their prayers, patience, understanding, support, the completion of this work would not have been possible. I want also to thank all the relatives in my homeland for their support and encouragement.

Finally, a very special thanks to my beloved wife, Suci Rimadheni, for her prayers and faithful support throughout the entire period of my PhD. My little ”kwa” and ”munga”, Nacitha Feiyazzarifa always “gave” me extra energy and “motivated” me to finish this thesis as soon as possible. I am truly indebted to their patience and understanding. Thank you!

To all those cited by name here and the many more I did not mentioned by name, I will forever owe you this achievement. Thank you all very much!

Abstract

Condition-based maintenance is an emerging paradigm of modern health monitoring, where maintenance operations are based upon diagnostics and prognostics. Prognostics promises to optimise maintenance scheduling, resources and supply chain management, leading to reductions in operational disruption, spares inventory, maintenance labour cost and hazardous conditions. The main objective of this research is to develop generic data-driven prognostic approaches to address several challenges associated with complex system prognostics, where in this particular work, the developed techniques are applied to the degradation data obtained from civil aerospace gas turbine engines.

This thesis contains four key contributions. Firstly, deterministic Bayesian prognostics is used to deal with large uncertainty in degradation data. The novelty and value in the presented formulation lies in a fuller Bayesian treatment of observation error than prior art while retaining the closed-form solution desirable for real-time, deterministic computation. Secondly, the Bayesian hierarchical model (BHM) is introduced to optimise the use of fleet data from multiple assets. This formulation allows Bayesian updates of an individual predictive model to be made, based upon data received from a fleet of assets with different in-service lives. The results obtained demonstrate BHM capability in dealing with some extreme scenarios, occurring in complex system prognostics. The next contribution lies in developing variational inference for the existing BHM to overcome the computational and convergence concerns that are raised by sampling methods needed for the inference of the original formulation. This technique delivers an approximate but deterministic solution, where the quality of approximation is found to be satisfactory with respect to prediction performance, computational speed and ease of use.

In the final contribution, an integration concept is proposed, combining the Bayesian data modelling technique with an information theoretic change-point detection algorithm to solve a wide class of prognostic problems, such as information arising from irregular events occurring during the life-cycle of an asset. This integration concept has a great potential to be implemented in complex system prognostics as it demonstrates several advantages of the deterministic BHM in combination with change-point detection to utilise, optimally, all available multiple unit data as well as data available at various levels of the system hierarchy.

Contents

List of Figures

List of Tables

General Abbreviations

ARIMA	Autoregressive Integrated Moving Average
BnHM	Bayesian non-Hierarchical Model
BHM	Bayesian Hierarchical Model
BR	Bayesian Regression
BR-1	Bayesian Regression 1
BR-2	Bayesian Regression 2
BR-3	Bayesian Regression 3
BR-3-VB	Bayesian Regression 3 using variational Bayes
BR-3-VB+CPD	BR-3-VB with CPD
BVAR	Bayesian Vector Autoregressive
CBM	Condition-Based Maintenance
CPD	Change-Point Detection
DLM	Dynamic Linear Model
EHM	Engine Health Monitoring
GP	Gaussian Process
iid	independent and identically distributed
Int-BR-3-VB	Integrated BR-3-VB
KM	Kaplan-Meier
MCMC	Markov Chain Monte Carlo
OLS	Ordinary Least Square
OSyS	Optimized Systems & Solutions
PHM	Prognostic Health Management
PM	Preventive Maintenance
RMS	Root Mean Square
RUL	Remaining Useful Life
RuLSIF	Relative unconstrained Least-Squares Importance Fitting
RVM	Relevance Vector Machine
SM	Scheduled Maintenance
TGT	Turbine Gas Temperature

General Symbols

cov	covariance
$\mathbb{E}$	Expectation
Gam	Gamma (distribution)
IG	Inverse Gamma (distribution)
KL	Kullback-Leiber divergence
$\mathcal{N}$	Normal (distribution)
NIG	Normal Inverse Gamma (distribution)
$\mathbf{\Phi}$	Design matrix
PE_α	α-relative Pearson divergence
St	Student-t (distribution)
var	variance
$\mathcal{W}$	Wishart (distribution)

Chapter 1

Introduction

1.1 Motivations

1.1.1 Condition-based maintenance

Today's complex and advanced systems require highly costly and sophisticated maintenance strategies. The expenditure of funds for maintenance of complex systems, such as structure, machinery, or equipment, amounts about half the initial investment cost before obsolescence forces replacement [Saunders, 2007]. An even more alarming fact is that one-third to one-half of this expenditure is wasted through ineffective maintenance [Heng et al., 2009]. This may be relatively small in comparison with the loss of production attributable to unwarranted in-service failures.

Energy efficiency is also related closely to maintenance strategy in that the condition of the system is reflected in how effectively the tasks are performed and how much energy it consumes. For example, if a car is not checked and maintained regularly, faults may occur which would cause car to consume more fuel which in-turn would not be environmentally friendly and result in an increase in its running costs. Another example is in gas turbine engine, where inappropriate maintenance would lead to a significant reduction in performance in which it gets progressively worse with the increasing of operating time [Naeem et al., 2001]. Any engine deterioration adversely affect the fuel consumption and life usage [Naeem et al., 1998]. This argument also applies to any system where effective maintenance means more efficient performance. Selecting proper maintenance strategy will improve the quality, efficiency and effectiveness in achieving

and enhancing company competitiveness [Alsyouf, 2004].

Figure 1.1 shows classification of maintenance strategy based on Eti et al. [2006]; Shenoy et al. [1997]. Traditionally, maintenance strategy is based on breakdown maintenance, where the experience of personnel familiar with equipment plays a part when equipment fails unexpectedly. However, this practice is not always sufficient when dealing with complex systems with numerous interrelating failure modes. Maintenance based on human expertise also becomes increasingly hard to deal with, due to ageing workers. There may be not enough new skilled workers to replace experienced and senior workforce who are leaving for retirement [Wireman, 2010]. Moreover, major causes of unscheduled maintenance events are unanticipated and extreme operating scenarios, which lead to serious operational issues, such as failure of missions and disruption costs. Analysis of maintenance costs has shown that a repair made after failure will normally be three times greater than the same repair made on a scheduled basis [Mobley, 2004].

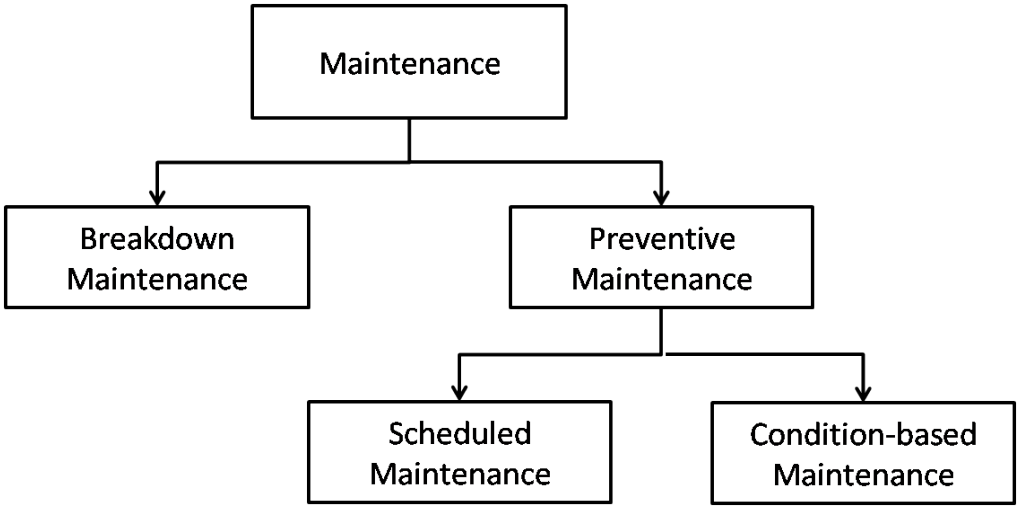


Figure 1.1: Classification of maintenance strategy.

To anticipate failures, an efficient repair of equipment before it fails is required by performing preventive maintenance (PM). PM is defined as removing a functioning device from operation in order to repair, test or inspect it prior to failure [Nachlas, 2005]. PM can be divided into two groups: scheduled maintenance (SM) and condition-based maintenance (CBM).

Conventional PM is based on SM, where the frequency of removals may be based

on age or usage through use of a conservative statistic, the so called "safe life removal interval". Statistical reliability distributions are based on the collective behaviour of a population of individuals acting in a specific environment. The reliability distribution must capture the spread of failure behaviour resulting from each individual having its own sources of durability variation caused by manufacturing, material or maintenance. However, Byington et al. [2004b] claimed that, based on historical evidence, the actual usage of components/systems, such as military aircraft, is often significantly different from the intended usage and operating environment. For example, usage will depend on the pilot and flying style in aviation systems. Even though SM reduces equipment failures, it is more labour intensive, may not eliminate catastrophic failures and cause unnecessary maintenance, e.g. over-maintenance and under-maintenance. Over-maintenance causes material waste and low equipment utilisation where, on the other hand, under-maintenance results in unexpected down events.

CBM becomes an alternative maintenance strategy to overcome the above issues. CBM is the process of collecting real-time sensor information from a functioning device (without interrupting normal machine operation) in order to reason about its health [Gebraeel et al., 2005; Heng et al., 2009]. Two important aspects of CBM programme include diagnostics (detection, isolation and identification of faults) and prognostics (the prediction of failure times) [Jardine et al., 2006].

A number of definitions of prognostics have been cited that seem to mix prognostic and diagnostic activities. For example Schwabacher and Goebel [2007] defined prognostics as detecting the precursors of a failure, and predicting how much time remains before a likely failure. It is suggested that this refers to both diagnostics and prognostics and a definition closer to that suggested by Saxena et al. [2008b] that prognostics is the estimation of remaining useful life (RUL), where RUL is defined as the time until the functional requirements can no longer be met. Sikorska et al. [2011] introduced a simple delineation regarding the relationship between diagnostics and prognostics: "diagnostics involves identifying and quantifying damage that has occurred, whilst prognostics are concerned with trying to predict the damage that is yet to occur". Therefore, prognostics can be considered as an extension to diagnostics and defined as: prediction, with quantified certainty, of future residual functional capability.

Prognostics promises to significantly reduce operational disruption, spares inventory, maintenance labour costs and hazardous conditions. However, prognostics is a

relatively new research area and has yet to receive its prominence compared to the other areas of CBM [Heng et al., 2009].

A considerable amount of prognostic research has been conducted to improve estimation of RUL of engineering assets. Examples include automotive [Abbas et al., 2007; Ompusunggu et al., 2012; Zanardelli et al., 2005] and expensive equipment in heavy industries, such as oil and gas [Panesar and Markeset, 2008; Zhan et al., 2011] and power plant [Bond et al., 2011]. Prognostics has also been widely used in various aerospace domains including avionics [Celaya et al., 2008; Hecht, 2006; Johansson and Leisner, 2012; Pecht and Jaai, 2010; Xu and Xu, 2011], unmanned aerial vehicles (UAV) [Walker, 2010], defence, such as the Joint Strike Fighter programme [Brown et al., 2007; Gao et al., 2008] and civil aerospace [King et al., 2009; Ling and Mahadevan, 2012; Rezaei and Dadouche, 2012]. In this work, the application is focused on prognostics for civil aerospace gas turbine engines.

1.1.2 Aerospace gas turbine engine prognostics

For many of the world's largest manufacturers, aftermarket service and parts operations essentially define the business. According to Dennis and Kambil [2003], after-sales services and parts contribute only 25% of revenues across all manufacturing companies but are responsible for 40-50% of profits. Table 1.1 shows that the aerospace and defence business accounts for about 47% percent of revenue, the largest in comparison to other global industries [Koudal, 2006].

Engine manufacturers, e.g. General Electric, Pratt & Whitney and Rolls-Royce, all have performance-based contracts with commercial airlines in which their compensation is tied to product availability (hours flown) [Kim et al., 2007; Marinai et al., 2004]. Services, such as TotalCare® and power by the hour arrangements, are now regarded as an essential element of delivering asset operation [King et al., 2009].

The economic impact of such service contracts is significant. For example, Rolls-Royce, one of the world's largest jet engine and gas turbine makers, has more than 14,000 aerospace engines in service, operated by more than 500 airlines and powering more than 5.5 million commercial flights per year [OSyS, 2013]. The considerable number of the engines to be maintained, in terms of service and providing proper spare-parts, enable this company to generate revenue about 55% of the more than US$11

Global industry	Share of service and parts business in overall sales
Aerospace and defence	47%
Automotive and commercial vehicle	37%
Diversified manufacturing and industrial products	20%
High technology and telecommunications equipments	19%
Life science/medical devices	21 %
All companies	26%

Table 1.1: Revenue impact of service and parts business by global industry [Koudal, 2006]

billion in total revenues [Rolls-Royce, 2013].

This evidence emphasises the significant benefits of applying health monitoring in civil aerospace gas turbine engines, which is one of the effective ways to reduce life cycle costs, improve engine reliability as well as availability [Li and Nilkitsaranont, 2009; Marinai et al., 2004]. Figure 1.2 illustrates the benefits of health monitoring in civil aerospace industry, based on Leao et al. [2008]. Aircrafts are highly valuable assets and large budgets are spent in aircraft support, maintenance and logistics. The application of prognostic technologies in civil aerospace can potentially yield profits to commercial aircraft operators.

Prognostics improves troubleshooting, enhances root cause analysis, and assists to prepare for maintenance in advance. Therefore, maintenance planning can be integrated, so that operation, material, logistics and human resources can be optimised to increase availability and reduce maintenance costs. An effective prognostic system is also capable of collecting and storing useful historical information, such as failure modes, operating conditions, environmental conditions and possible deficiencies. System designers can improve reliability by redesigning of critical components and subsystems with this information [Sun et al., 2012]. This would improve systems performance, reduce energy consumption and useful lifetimes can be extended.

This thesis focuses on the development of prognostic algorithms to estimate RUL

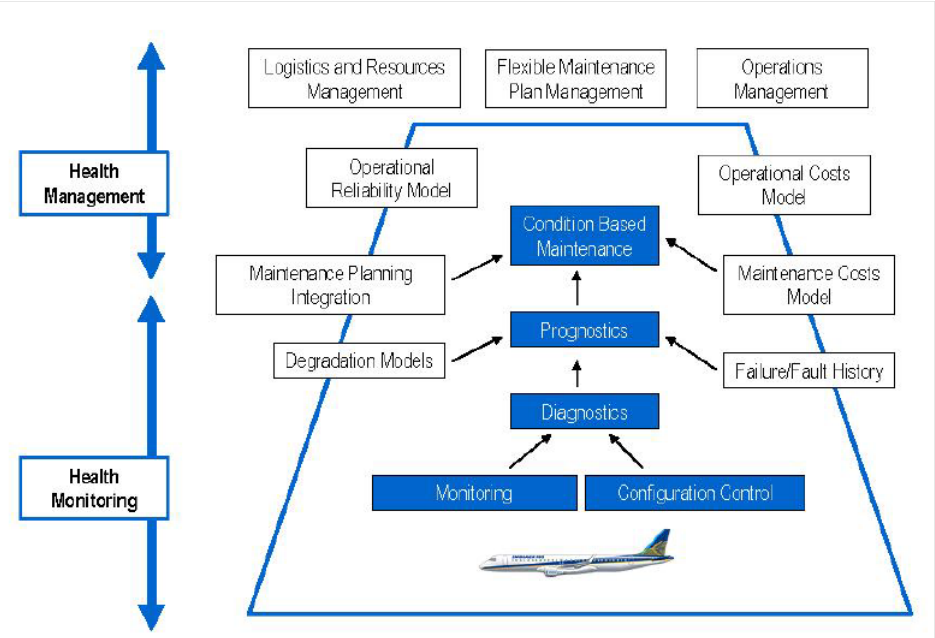


Figure 1.2: Health monitoring benefits in civil aerospace industry [Leao et al., 2008]. Prognostics optimises maintenance planning integration to enhance aircraft availability and reduces maintenance cost.

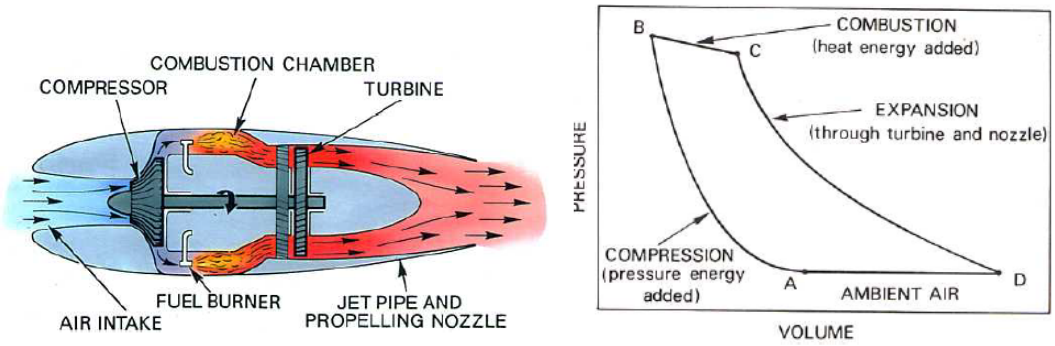
of complex systems in general and demonstrates their efficacy on civil aerospace gas turbine engines.

1.2 Problem Definition

This section provides a problem definition in the context of prognostic health monitoring for complex systems, such as gas turbine engines. First, a brief description of the principle of gas turbine engines is provided. Subsequently, gas turbine engine degradation is described, including the degradation factors and the health parameter. Finally, several challenges associated with this class of complex systems are discussed.

1.2.1 Gas turbine engine principles

The concept of the gas turbine has been acknowledged to an English coalmaster and inventor, named John Barber (1734-1801), who patented his idea about gas turbine in



(a) (b)

Figure 1.3: A Whittle-type turbo-jet engine schematic and its working cycle [Rolls-Royce, 1996].

1791 [Davey, 2003]. He established the basic principle of a gas turbine engine, despite the lack of technology at that time [Hill and Peterson, 1992]. Material, design and manufacturing techniques needed to put this principle into a working machine were not fully available until the early parts of the 20^{th} century. The first patent was granted to Frank Whittle (1907-1996) in 1930 for using a gas turbine to produce a propulsive jet [Rolls-Royce, 1996].

Gas turbine engines are widely used in different fields to generate energy. They are also commonly used in aircraft. A gas turbine engine is essentially a power plant which utilises air as a working fluid to produce power in the form of thrust, shaft-power or compressed air. Figure 1.3a shows a Whittle-type turbo-jet engine [Rolls-Royce, 1996]. It can be seen that there are several internal sections inside, including inlet section, compressor section, combustor section and turbine section.

In addition, figure 1.3b illustrates the working cycle on a pressure-volume diagram. The engine cycles show that in each instance there is induction, compression, combustion and exhaust. In the inlet section, the air intake and fan directs the air into the engine (point A), then the compressor compresses the air to a high pressure (point A to B). After the compressor section, the high-pressure air is directed into the combustion chamber, where fuel (kerosene) is spread as small particles and burned at high temperature and constant pressure, thereby considerably increasing the volume and velocity of air (point B to C). The high velocity air is then directed towards the turbine and driving it using the kinetic energy from the high-speed gas. A portion of the high-velocity air is

expanded through the exit nozzle, producing thrust (point C to D) [Rolls-Royce, 1996].

When gas turbine engines are run, they become fouled with airborne contaminants such as oil, pollen, soot, unburned fuel, soils and salt which encrust compressor components [Kurz and Brun, 2007]. Therefore, gas turbine engines show the effects of damage and deterioration in its lifetime of service. The degradation of an engine has an adverse effect on the engine's overall performance [Khani et al., 2012]. The following section will discuss how to determine the main parameter involved in prognostics.

1.2.2 Gas turbine engine degradation

Various factors affect degradation in gas turbine engine performance, including [General-Electric, 2008; Kurz and Brun, 2001; Malinge and Courtenay, 2007]:

- dust/dirt ingestion and further accumulation on fan blades/compressor airfoils
- increased air seal, compressor and turbine blade-tip clearances because of rub
- other mechanisms such as erosion of airfoils and seals, hot section oxidation
- foreign object damage

To reduce maintenance cost and avoid service disruption, engine health monitoring (EHM) has been employed in modern gas turbine engines. EHM is a pro-active technique for predicting when something might go wrong (prognostics) and preventing a potential threat before it has a chance to develop into a real problem, e.g. fault. EHM can be used to estimate the health of thousands of engines operating worldwide, using on-board sensors and live satellite feeds [Waters, 2009].

There are several sensors fitted to monitor critical engine characteristics, such as temperatures, pressures, speeds, flows and vibration levels, to ensure they are within acceptable tolerances and to highlight when they are not. Figure 1.4 shows the typical parameters measured on the Rolls-Royce Trent 900 engine [Rolls-Royce, 2013].

From several aforementioned parameters, the main challenge is to determine properly a generic health index[1], which can represent the overall health of a gas turbine engine. As the engine efficiency reduces, the fuel supplied will necessarily increase to

[1]The main health feature to be forecast.

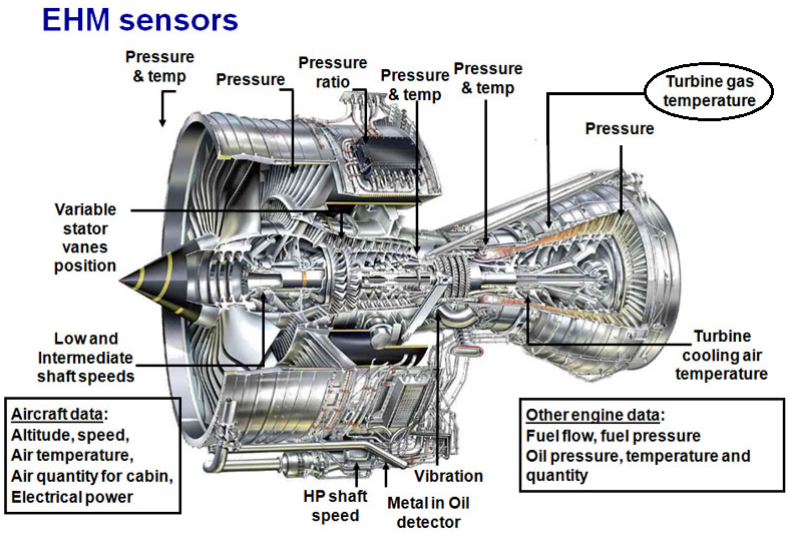


Figure 1.4: The location of the EHM sensors on the Rolls-Royce Trent 900 engine schematic [Rolls-Royce, 2013]. Turbine Gas Temperature (TGT) is measured at the turbine exit (oval region)

generate the required thrust. This results in an increase of the temperature in the engine. Therefore, the global health of the engine can be derived from the core flow temperature, measured at the turbine exit. This is called Turbine Gas Temperature (TGT).

TGT is a critical variable for engine operation and it is essential to provide an indication of this temperature. Ideally, the measurement should be made at turbine entry temperature. However, this is not practical, due to the high temperatures in that section. As the temperature drop across the turbine varies in a known manner, the temperature at the outlet from the turbine is usually measured by suitably positioned thermocouples [Rolls-Royce, 1996]. As shown in figure 1.4, the oval region shows where TGT is measured on the Rolls-Royce Trent engine [Rolls-Royce, 2013].

An estimate of the difference between the certified TGT (operational limit) and a projection of TGT to full-rated take-off at reference conditions is named TGT margin [Malinge and Courtenay, 2007]. The TGT margin is usually used to monitor the gas path degradation of the engine, to detect the changes in performance for each engine, and to indicate the need for inspection/maintenance. Hence, TGT margin can be used to forecast RUL [Malinge and Courtenay, 2007; Marinai et al., 2003; Müller et al., 2010; Walsh and Fletcher, 2004]. Figure 1.5 shows an example of normalised TGT margin of

gas turbine engine degradation.

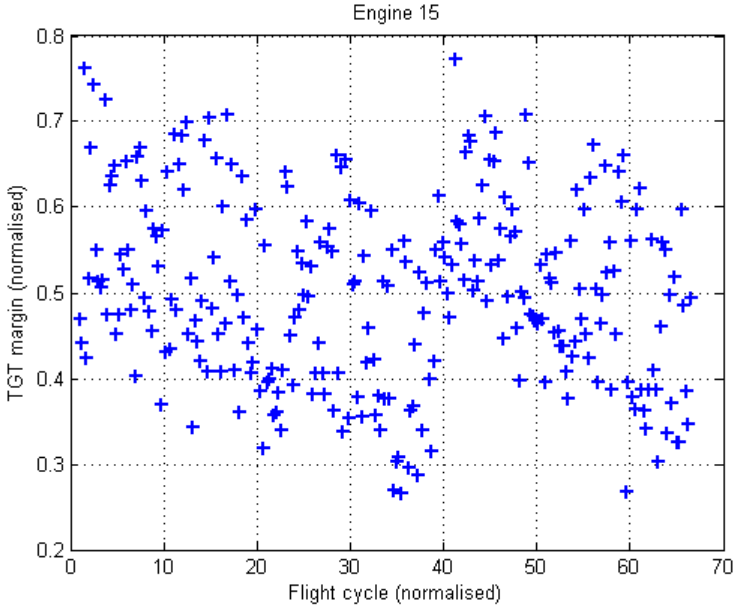


Figure 1.5: Example of normalised TGT margin data. The maintenance effect can be seen around flight cycle at 40 where the TGT margin degradation has recovered.

In this work, TGT margin data, obtained from Rolls-Royce Trent 500 engines, is used for validation of the developed prognostic algorithms. This data is supplied by Optimized Systems & Solutions (OSyS)[1]. All data, shown in this thesis, has been normalised, for confidentiality.

1.2.3 Challenges in gas turbine engine prognostics

Prognostic algorithms play a crucial role in RUL estimation of a complex system. For successful practical implementation, selecting a suitable prognostic algorithm requires a good understanding of the challenges associated with a specific application [Sikorska et al., 2011]. This section describes various challenges associated with a class of complex systems in particular gas turbine engine.

[1]Optimized Systems & Solutions Limited (OSyS) is a part of the Rolls-Royce group and provides asset optimisation solutions [OSyS, 2013].

1.2.3.1 Selecting algorithm for complex systems

A complex engineering system is defined as a group of interrelated, interacting or interdependent constituents (components) forming a complex whole [Jamshidi, 2010]. Many engineering systems comprise hundreds or thousands of components. Intermediate groupings, or various levels of subsystems, are necessary to describe or depict these systems correctly.

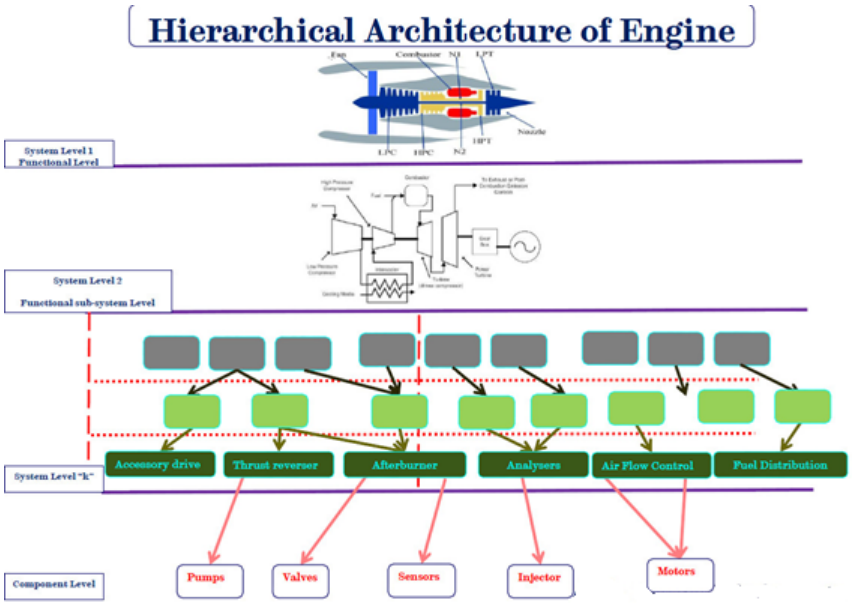
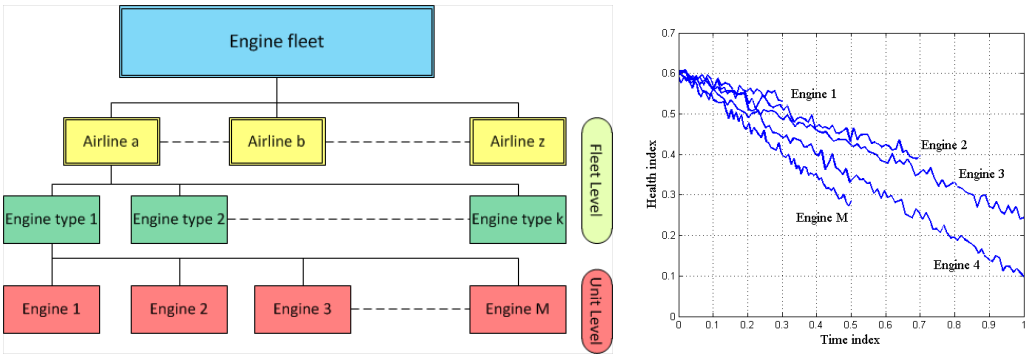


Figure 1.6: Hierarchical system of aerospace gas turbine engine where it comprises multiple levels, such as subsystem and component levels. Engine degradation takes place on the system level may also be influenced by the changes which occur at other levels.

Such an engineering system that requires one or more levels of definition intermediate to system and component is characterised as a complex system in this study. Thus, a complex system is a system composed of a number of subsystems, each of which is embodied by a particular set of components, or sub-subsystems. Hence, an aircraft engine can be considered as a complex system comprising multiple interacting subsystems. Engine degradation, which takes place on the system level (e.g. TGT margin), may also be influenced by the changes that occur at subsystem or even component level. In figure 1.6, an aircraft engine is presented as an example of a complex system, which comprises subsystems, such as Low Pressure Compressor (LPC), High Pressure Turbine



(a) (b)

Figure 1.7: A hierarchy of the engine fleet across airlines, engine type and individual engine (left) and multiple degradation data available (right). A considerable volume of health signal data generated from the fleet of engines is advantageous for data-driven prognostic algorithm.

(HPT). Further down the hierarchy, the subsystems are composed of components (e.g., HPT blades).

Therefore, it is difficult to construct a model based on physical principles that mimics the dynamics of the system's long-term degradation. Modern gas turbine engines are already fitted with numerous sensors for control and monitoring purposes [Waters, 2009]. Furthermore, the fleet of engines, illustrated in figure 1.7, is capable of generating a considerable volume of health signal data. This motivates the direct use of in-flight data to achieve the prognostic goal. It can be concluded that a data-driven approach is promising for the RUL estimation for civil aerospace gas turbine engines.

1.2.3.2 Inexistence of ground truth

Another complication that arises is the inexistence of ground truth, i.e. assets are never allowed to fail in service, so true RUL can only be estimated. Therefore, an appropriate methodology is required for testing and validating prognostic algorithms. In this work, synthetic degradation data of known properties is generated to test the algorithms. The data are generated adequately to emulate important characteristics of the real degradation data. This methodology will be applied throughout this thesis, prior to validation with real data.

1.2.3.3 Dealing with uncertainty

As shown in figure 1.5, there is large uncertainty associated with TGT margin data as it gets corrupted with noise owing to gas turbine design, manufacturing, ambient and environmental condition, operating condition, duty mission, maintenance action, etc [Li and Nilkitsaranont, 2009].

Large uncertainty in the data may cause inconsistency in prognostic prediction, especially when there is little data available. In other words, an irrational prediction may arise, e.g. prediction may show improving health. A suitable class of algorithm should be selected, where it should have the capability to enable variation and uncertainty to be quantified, mainly by using distributions instead of point estimates in risk assessment.

1.2.3.4 Accommodating heterogeneous fleet

Another question is: how can a prognostic algorithm use, optimally, the data available from a fleet of engine (figure 1.7) when estimating the RUL of a specific engine (data sharing).

A sophisticated approach should be developed for accommodating the heterogeneous fleet into a single prognostic model. Therefore, it should have the ability to use the measures of degradation from one asset or asset type and apply this information to inform the prediction for other units, i.e. the information sharing capability, to enhance RUL estimation for a particular unit.

1.2.3.5 Handling multiple degradation patterns

In practice, research reveals that there are many patterns of failure which actually occur in engineering assets [Moubray, 1997]. In gas turbine engine degradation, most of the observed patterns of degradation are nearly linear [Li and Nilkitsaranont, 2009; Puggina and Venturini, 2012], however there are cases where the rates of degradation may be non-linear [Li and Nilkitsaranont, 2009; Saravanamuttoo et al., 2009], where degradation may be at an approximately constant rate for a period of time followed by an increase in rate. The latter is caused by various factors, including a step change in

covariates[1] and fault modes[2] [Ackert, 2010].

To overcome this issue, an effective approach is needed, where it should be able to detect and analyse the sources of non-linearity, to then provide information to a prognostic algorithm to update its belief about degradation.

1.2.3.6 Coping with recoverable system

A gas turbine engine is considered as a recoverable system which its performance can be improved by a maintenance action. This condition is illustrated in figure 1.5, where TGT margin increases at approximately flight cycle 40.

The gas turbine engine becomes fouled with airborne contaminants such as unburned fuel, oil, solids and pollen which encrust compressor components. Proper operation and maintenance can be used to minimise the fouling type losses. For example, compressor washing is an effective method to maintain the compressor efficiency by removing fouling. The washing of gas turbine compressors maximises the power output, and fuel efficiency as well as, increasing the life time of the compressor components [General-Electric, 2008; Kurz and Brun, 2001; Malinge and Courtenay, 2007].

For accurate prognostics, knowledge of maintenance actions which affect the rate and state of degradation is important. However, this information is often difficult to obtain and incorporate [Brasco et al., 2013], because the maintenance actions are performed at geographically dispersed locations by organisations independent to those performing fleet management. These factors lead to uncertainty in the maintenance state of the asset. The uncertainty surrounding maintenance actions impacts the ability to accurately predict and extrapolate the degradation of a unit [Skaf et al., 2013].

An appropriate solution is to detect accurately maintenance events, directly from the measured service data. The knowledge about the schedule of maintenance events assist in predicting the future behaviour of the asset. The prognostic algorithm can be reset, to then restart the prediction.

[1] Covariates are any factors which affect degradation, such as operating conditions and environmental effects.

[2] Fault modes are specific types of fault. Buckling, creep, fatigue, corrosion, and wear are examples of mechanical fault modes.

1.3 Outline of thesis and contribution

The main objective of this research is to develop generic data-driven prognostic algorithms with applicability in various scientific and engineering fields, where in this particular work the developed techniques are applied to data obtained from gas turbine engines. The developed algorithms aim to cope with various challenges, described in section 1.2.3. In this research, all developed algorithms are written and implemented in MATLAB.

The subsections below list the major contributions of the thesis as well as provide a description about the contents of the thesis.

Chapter 2

Description: This chapter provides a literature review, covers a wide perspective of prognostics, including prognostic health management (PHM). Available prognostic techniques from a variety of research and application disciplines are reviewed providing the methodology of algorithm selection. Finally, this chapter describes the principle of Bayesian approach as selected prognostic algorithm.

Contribution: Prognostic algorithms are classified and discussed, providing the motivations for selecting data-driven Bayesian approaches. The principle and potential of Bayesian prognostics are highlighted. This literature review, including a generic Bayesian prognostic method, was presented in *The Eight International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Cardiff, UK, June 2011.

Chapter 3

Description: This chapter describes a Bayesian prognostic technique, combining two sources of information: historical in-service data from the engine fleet population and once-per-flight transmitted performance measurements. Using this data, the Bayesian technique presents predictive results within well defined uncertainty bounds.

Contribution: The novelty and value in the presented Bayesian formulation lies in a fuller Bayesian treatment of observation error than prior-art while retaining the closed-form solution desirable for real-time, deterministic computation. This approach shows improvement in terms of accuracy and consistency in dealing with high level of noise in degradation data. This contribution[1] was presented in *IEEE Aerospace Conference*, Big Sky, Montana, USA, March 2013.

Chapter 4

Description: This chapter introduces a prognostic technique using a Bayesian Hierarchical Model (BHM), which utilises fleet data from multiple assets to perform probabilistic estimation of RUL for gas turbine engines. The hierarchical formulation allows Bayesian updates of an individual predictive model to be made, based upon data received from a fleet of assets with different in-service lives.

Contribution: First contribution is in simplifying the complexity of prior specification for BHM, by introducing a methodology to produce "moderately" informative prior distribution based on in-service data. Second contribution lies in the implementation of BHM prognostics, which is appropriate to maximise the use of an asset's data by accommodating them into a hierarchical model. In this way, health information can be shared between the engines in order to enhance the RUL estimation, especially in extreme prognostic scenarios, such as the change in the slope and very high noise problems. Furthermore, this technique is also able to predict, simultaneously, degradation signals with different sample sizes, owing to the nature of the model. In order to estimate the RUL of a new engine, the degradation signal is simply fed into the model without the need to re-estimate new prior parameters. This work has been submitted to *Journal of Expert Systems With Applications*.

Chapter 5

Description: In this chapter, approximate inference based on variational Bayes is applied to the existing BHM, described in chapter 4, to overcome the computational,

[1] This paper has been chosen to be nominated for the best Rolls-Royce University Technology Centre (UTC) paper competition across all UTCs.

convergence and usability concerns that are raised by the numerical sampling techniques, needed for the inference of the original formulation. BHM using variational inference delivers an approximate but deterministic solution. The developed variational Bayesian prognostic algorithm is tested on gas turbine engine degradation data, where the quality of approximation is shown to be satisfactory with respect to prediction performance, computational speed and ease of use.

Contribution: This chapter delivers two main contributions. The first is in the derivation of solutions for the approximated posterior and predictive distributions as well as the lower bound for BHM prognostics, based on variational inference. The derivation is described in detail in appendix D. Another contribution lies in variational Bayesian prognostics, where it demonstrates successfully its capability to substitute Gibbs sampler for BHM prognostics. The achievement has been submitted to *Journal of Mechanical Systems and Signal Processing*.

Chapter 6

Description: This chapter presents an integration concept combining BHM with an information theoretic change point detection (CPD) algorithm to solve a wide class of prognostic issues, such as information arising from irregular events occurring during the life cycle of an asset, rapid degradation in the state of health parameter, etc. This concept is applied to two case studies. In the first case study, the CPD algorithm is used to detect the change in a covariate, which affects the slope of degradation and this information is then utilised to reconfigure the prognostic model to enhance the quality of RUL estimation. Furthermore, in the second case study, CPD is applied directly to the degradation data to discover when an unknown maintenance event took place. This information directs the main prognostic algorithm to reset the prediction. The results for both case studies reveal a significant improvement in quality of RUL estimation.

Contribution: The proposed integrated prognostic concept proves to be promising and demonstrates several advantages of the hierarchical model in combination with CPD algorithm to improve utilisation of the available multiple engine data as well as data available at various levels of a system's hierarchy. This method is able to detect

changes or faults in multiple covariates (e.g. vibration, ambient temperature) at any level of the system to then informing main prognostic algorithm to update belief about degradation. The recovery in degradation data due to maintenance action can also be handled automatically based on this concept. Part of this work was presented in *Annual conference of the prognostics and health management society*, New Orleans, USA, October 2013, whereas the whole contribution of this chapter has been submitted to *Journal of Control Engineering Practice*.

Chapter 7

This chapter states the general conclusions of this thesis and outlines future work.

1.4 Publications

As described in section 1.3, several contributions have been provided in several publications. To summarise the contributions, this section provides a list of the publications.

1. **M.A. Zaidan**, R. Relan, A.R. Mills, and R.F. Harrison. Integrated Bayesian prognostics of complex systems. *Control Engineering Practice*, submitted 2014.

2. **M.A. Zaidan**, A.R. Mills, R.F. Harrison and P.J Fleming. Gas turbine engine prognostics using Bayesian hierarchical models: a variational approach. *Mechanical Systems and Signal Processing*, submitted 2014.

3. **M.A. Zaidan**, R.F. Harrison, A.R. Mills, and P.J. Fleming. Bayesian hierarchical models for aerospace gas turbine engine prognostics. *Journal of Expert Systems with Applications*, submitted 2014.

4. O.W. Laslett, A.R. Mills, **M.A. Zaidan**, and R.F. Harrison. Fusing an ensemble of diverse prognostic life predictions. In *IEEE Aerospace Conference*, Big Sky, USA, March 2014.

5. Z. Skaf, **M.A. Zaidan**, R.F. Harrison, and A.R. Mills. Accommodating repair actions into gas turbine prognostics. In *Annual conference of the prognostics and health management society*, New Orleans, USA, October 2013.

6. **M.A. Zaidan**, A.R. Mills, and R.F. Harrison. Bayesian framework for aerospace gas turbine engine prognostics. In *IEEE Aerospace Conference*, Big Sky, USA, March 2013.

7. **M.A. Zaidan**, A.R. Mills, and R.F. Harrison. Towards enhanced prognostics with advanced data-driven modelling. In *The Eighth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, Cardiff, UK, June 2011.

Chapter 2

Literature Review

2.1 Introduction

This chapter describes a literature review to provide a broad view of prognostics, including prognostic health management (PHM) and prognostic algorithms. Available prognostic techniques are explored from a variety of research and application disciplines. The primary objectives are to present an overview of various prognostic techniques and to provide a methodology: how the most appropriate approach for gas turbine engine prognostics is selected.

2.2 Prognostic health management

As an engineering discipline, PHM is a technology to provide users with an integrated view of the health state of a machine or an overall system, which aims to enhance its reliability as well as its availability in its life-cycle by detection of current and approaching failures and by providing for mitigation of the system risks [Lee et al., 2013]. PHM incorporates hardware, software, modelling and analysis in support of prediction. For example, hardware (e.g. sensor system) is required to make current measurements of key parameters in the system. Sensor system, data pre-processing, knowledge elicitation and mathematical techniques are some important elements in PHM. Several factors influence the performance of prognostic tasks: each process step has specific functions and contributions to overall prognostic accuracy. This section provides a wider view of

the factors that influence prognostic accuracy.

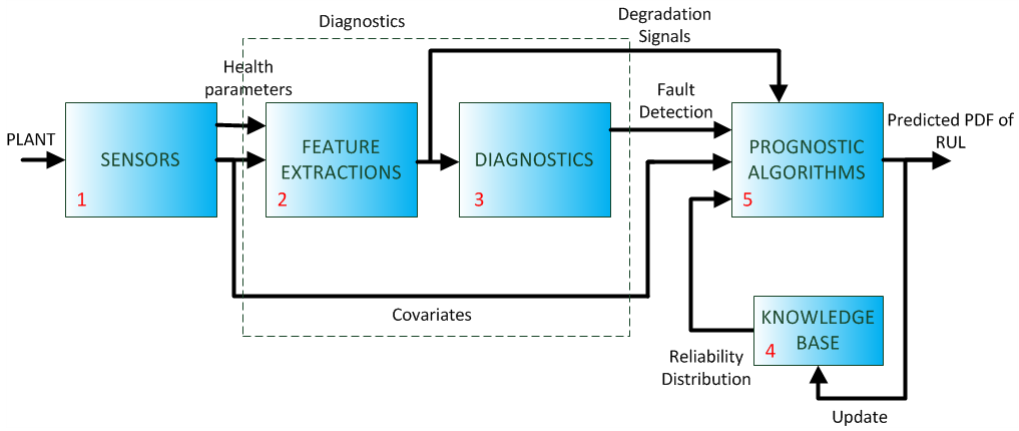


Figure 2.1: Diagram of the process of PHM. PHM considers all these elements for the success of overall prognostics programme.

Figure 2.1 illustrates a diagram of the PHM process. In the first block, sensing and data acquisition are major issues that must be addressed when developing prognostic solutions. The first challenge in a sensing system for prognostics is to determine the types of data to be acquired. The data types can be classified into three general categories. The first type is kinematic quantities. For example, accelerometers to measure the vibration, and micro-electro-mechanical systems (MEMS) to measure strain, rotational and linear acceleration. The second and third types are environmental quantities (e.g. temperature, pressure, moisture, etc) and operational quantities (e.g. load, speed, amount of fuel, etc), respectively [Farrar et al., 2005]. In other words, the sensors are utilised to monitor degradation signals (health index) as well as covariates[1]. In the case of a gas turbine engine, the main health index, TGT, is usually measured by suitably positioned thermocouples [Rolls-Royce, 1996].

Another major challenge in sensor system is to determine sensor properties, which need to be defined earlier and typically cannot be changed easily once the sensor system has been installed. It is vital to determine the number of sensors and the sensor locations for ensuring optimal sensing system. For example, the optimal location and the typical parameters measured by EHM sensors in a Rolls-Royce engine are shown in figure 1.4. Other important sensor properties include bandwidth, sensitivity (dynamic range),

[1]Covariates are explanatory variables which affect the degradation signal, such as health features, operating conditions and environmental factors.

stability, precision, resolution, reliability, power requirement, cost, etc [Farrar et al., 2005; Pecht, 2008; Powrie and Fisher, 1999].

Diagnostics includes feature extraction[1] and diagnostic decision (e.g. fault detection), shown in the second and the third blocks, respectively. Selecting a proper feature as a health index is crucial for the success of prognostic goal. For example, to characterise the health condition of rolling element bearings, Gebraeel et al. [2005] and Widodo and Yang [2011b] computed the root mean square (RMS) value and kurtosis of vibration signal, respectively. These health features are used to forecast the RUL of the rolling element bearings.

For mechanical systems, Lebold et al. [2000] and Qiu et al. [2003] discussed several feature extraction techniques for vibration signal monitoring. In material application, Ray and Tangirala [1996] used crack in aluminium alloy, whereas in electronic application, Saha and Goebel [2008] used capacities of battery cells, as their health indexes. In our case, as described in chapter 1.2.2, TGT margin is considered as a main health index, representing the overall health of a gas turbine engine. In order to estimate its RUL, TGT margin is forecast until it reaches maintenance threshold[2].

In addition, other covariates or diagnostic output may be used to support the prognostic algorithm in order to enhance RUL estimation. The tasks are to: detect abnormalities, locate the fault and identify failure modes [Sikorska et al., 2011; Yu et al., 2012]. Detecting (diagnosing) any changes (abnormality) in covariates is used to revise the trajectory prediction. For examples, Gebraeel and Pan [2008] used load and speed changes, to improve RUL estimation of bearings, whereas Ramakrishnan and Pecht [2003] used temperature and vibration data to estimate consumed life of an electronic component-board assembly placed under the hood of an automobile and subjected to normal driving conditions based on physics of failure. Other examples of incorporating covariates in prognostic approaches can be found in Heng et al. [2009]; Pecht and Gu [2009]. For the gas turbine engine case, Müller et al. [2010] described the effects of varying environmental (e.g. temperature, pressure, humidity and concentration of particles or other contaminants, like sulphur in the ambient air) and operating conditions (e.g. flight length) on engine deterioration. For our problem, the inclusion of event detection (diagnostics) will be discussed in chapter 6.

[1] Here, feature extraction includes raw data or processed data as health index.

[2] The maintenance threshold is typically defined by experts.

The knowledge base (the fourth block) contains prior information about the presupposed reliability and failure behaviour of components/systems. Physics of failure analysis, expert opinion, previous qualification tests and in-service data are examples of sources for the knowledge base. The first two of these sources can be used for model-based approaches whereas the last two can be employed for data-driven prognostics. In-service collected data offers potential to account for real component and environmental variability because identifying an actual operational relationship can be difficult from acceleration test data [Hamada et al., 2008].

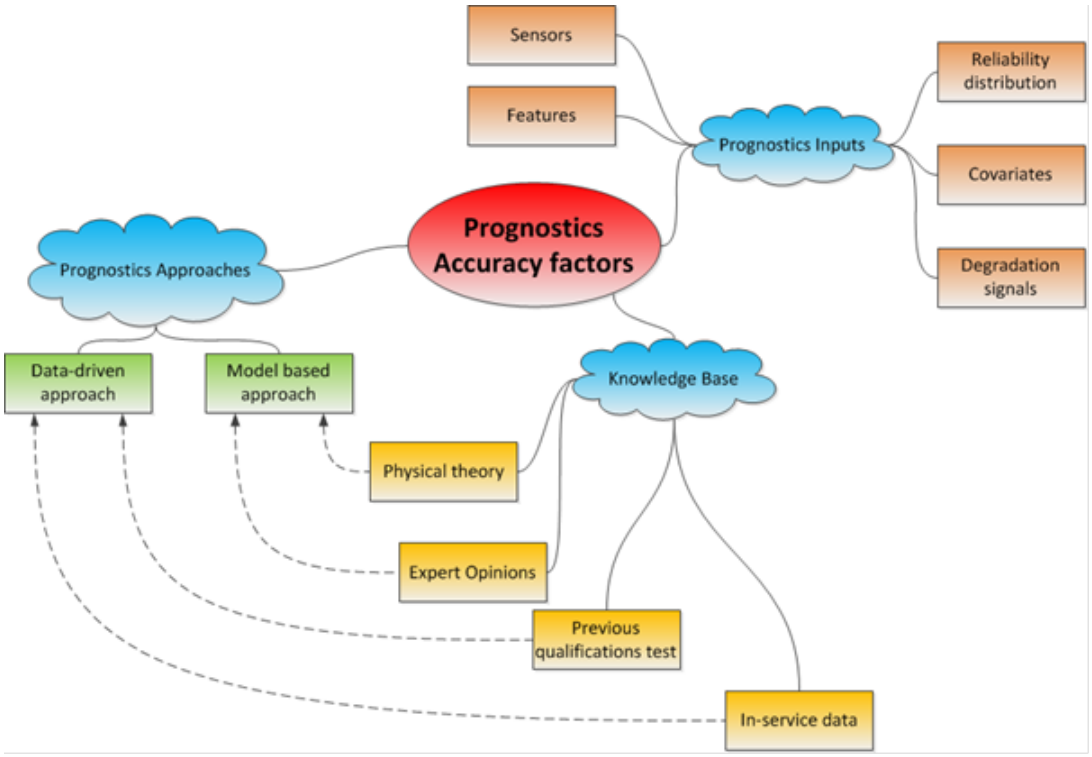


Figure 2.2: An illustration of various factors influencing prognostic performance, where this research focuses in prognostic algorithms.

Finally, prognostic algorithms play a very important task for the success of the prognostic goal [Schwabacher and Goebel, 2007], shown by the fifth block. Since, this research focuses mainly on the development of prognostic algorithms, a specific review in prognostic algorithms will be carried out in section 2.3.

In summary, each process in prognostics has a different function and affects the accuracy and performance of the overall system. This illustration is shown in figure

2.2, where inputs, methods and the knowledge base are key determinants of prognostic accuracy.

2.3 Prognostic algorithms

Developing prognostic techniques is an active research area [Lee et al., 2013; Saxena et al., 2008b]. Hundreds of papers in this area, including theory and practical applications, appear every year in academic journals, conference proceedings and technical reports. Despite the number of prognostic literatures is still much smaller than diagnostic works [Si et al., 2011], several survey papers on prognostic algorithms have been produced to simplify the selection of prognostic algorithms for researchers and practitioners.

Katipamula and Brambley [2005] and Jardine et al. [2006] carried out some simple reviews on the current status of prognostics, but these emphasise more in diagnostics. In addition, Schwabacher and Goebel [2007] surveyed data-driven prognostic approaches and classified them into two main categories: model-based and data-driven approaches, where they have made their own definitions about classical artificial intelligence, conventional numerical methods and machine learning. This classification may be controversial because it is difficult to distinguish between classical artificial intelligence and machine learning as well as conventional numerical method.

Heng et al. [2009] provided a prognostic survey, which includes state-of-art, challenges and opportunities. However, it focuses only on applications to rotating machinery. Sikorska et al. [2011] discussed business issues that need to be considered when selecting an appropriate prognostic method for trial, which also explores the strengths and weaknesses of each approach. They also provided classification tables and process flow diagrams to assist the users in industry and researchers in selecting appropriate prognostic models for predicting the RUL of engineering assets within their specific business environment.

Most recently, Lee et al. [2013] reviewed various methodologies and techniques in PHM research. The survey is not limited on prognostics, which is just one facet of PHM, but also includes diagnostics. They also presented a systematic methodology for conducting PHM as applied to machinery maintenance.

Other prognostic reviews have been presented in Goh et al. [2006]; Kothamasu et al.

[2009]; Peng et al. [2010]; Zhang et al. [2006]. These survey papers have provided an appropriately broad view of various prognostic techniques as well as indicating how the users and researchers can utilise and improve those approaches.

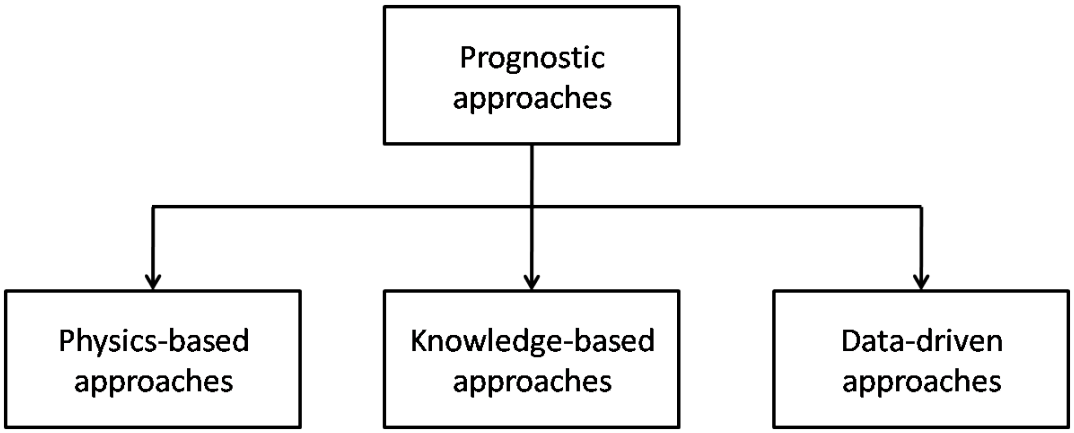


Figure 2.3: Classification of prognostic approaches.

In this review, prognostic approaches are divided into three categories: physics-based, knowledge-based and data-driven approaches, as shown in figure 2.3. This section does not aim to provide a review in detail as the above mentioned papers. Instead, this attempts to summarise and review briefly various prognostic approaches (with some examples) to provide a clear understanding of the methodology in selecting an appropriate prognostic approach for complex system applications.

2.3.1 Physics-based approaches

Physics-based approaches consider the physical processes and interactions between components in a system. Comprehensive mathematical representations, such as differential equations, are used to represent the system and physics-of-failure. The methods require specific knowledge and theories relevant to a particular monitored system. This type of approach is promising because the physics-based approaches consider environmental conditions such as humidity, vibration, shock and load, therefore it is possible to properly perform diagnostics and estimate RUL for known failure mechanisms. Furthermore, when the understanding of system degradation improves, these models can also enhance their accuracy [Peng et al., 2010].

The examples include a defect propagation model by mechanistic modelling for RUL estimation of bearings [Li et al., 1999], a stiffness-based prognostic model for bearing systems based on vibration response analysis and damage mechanics [Qiu et al., 2002] and Paris' law to predict RUL of a gear with fatigue crack growth [Li and Lee, 2005].

In addition, physics-based approaches have been widely utilised in electronic applications. Patil et al. [2009] used this type of model for prediction of the RUL of insulated gate bipolar transistors (IGBT) and Goodman [2001] developed prognostic methodology based on failure modes in semiconductor. Ramakrishnan and Pecht [2003] explored electronic failures subjected to thermo-mechanical loads. By monitoring the environment of a device over its life cycle, it may be possible to determine the amount of damage induced by various loads and to predict RUL accurately.

However, this type of approach also suffers from several drawbacks. Firstly, models are expensive to develop, because the specific domain experts need to be involved and the parameters in the model must be validated by large sets of data. This type of approach is also component specific, which means it cannot be applied to other types of components [Brotherton et al., 2000]. The final limitation is that most prognostic problems that can be solved by the physics-based approaches are at component level or subsystem level.

As described in section 1.2.3.1, for prognostics at system level[1], especially for a complex system, e.g. a gas turbine engine, it is difficult to construct a physics-based model that mimics the dynamics of the system's long-term degradation.

2.3.2 Knowledge-based approaches

Knowledge-based approaches evaluate the similarity between an observed situation and a database of previously defined failures and conclude the life expectancy from previous event [Sikorska et al., 2011]. Two major examples are expert systems and fuzzy logic [Peng et al., 2010].

[1] System level meaning overall system in general.

2.3.2.1 Expert systems

An Expert system is a software program which exhibits human knowledge in solving a specific domain problem. It typically comprises a knowledge base, obtained from subject matter experts, and a rule-base, for implementing that knowledge to solve a particular problem. Precise rules are normally formulated in the form IF condition, THEN consequence, based on heuristic facts acquired by one or more experts over a number of years. These rules can be specific domain rules or heuristic rules (rules of thumb) and can be combined together using logical operators [Garga et al., 2001].

For instance, Butler [1996] developed an expert system for incipient failure detection and predictive maintenance (FDPM) system for application to distribution systems. The FDPM system consists of an expert system engine, a knowledge base, mathematical and neural network models of aging of distribution equipment and historical measurements databases. This system is used to assess the integrity of a power distribution system component and predict the maintenance needs. Another example of expert system is PROMISE (PROgnostic and Intelligent Monitoring Expert System), developed by Biagetti and Sciubba [2004]. This expert system concentrates on plant shutdowns due to sudden failures by generating real-time information about the existence of faults, predicting the faults and providing suggestions about how to control the problem.

The advantages of this method are that outputs are understandable and reasoning for a particular result can be established. However, it is not always straightforward to obtain domain knowledge and convert it to rules, especially when the system complexity increases [Peng et al., 2010]. Another drawback is that expert systems cannot deal with new situations which are not covered explicitly in its knowledge bases.

2.3.2.2 Fuzzy Logic

Fuzzy logic uses linguistic variables to provide a human-like and intuitive way of representing and reasoning with incomplete and inaccurate information. A fuzzy system consists of a knowledge base, fuzzy rule base and algorithms for applying the logic.

Fuzzy logic was used in a chemical pulp mill for real-time process condition monitoring and incident prevention [Feng et al., 1998]. In addition, this method was applied for diagnostics and prognostics of bearing faults in induction motors [Satish and Sarma, 2005].

Furthermore, Majidian and Saidi [2007] applied fuzzy logic to estimate the remaining life of tubes in a power boiler and the results were then compared with neural network. They concluded that neural networks for life prediction is easier in comparison with fuzzy logic because membership function in fuzzy logic was defined through human knowledge.

Despite fuzzy logic becomes an effective method when there is incomplete/inaccurate data (which is commonly found in practice) or no mathematical model is available or implementable, this method relies on the availability of a suitable expert to specify the rules underlying system behaviour and develop the fuzzy sets representing each variable's characteristics [Sikorska et al., 2011]. This method's requirement becomes more difficult when the experts need to specify rules for a complex system with numerous of interrelating failure modes.

2.3.3 Data-driven approaches

Data-driven approaches attempt to model system behaviour directly from historical data instead of building models based on comprehensive system physics and human knowledge. These methods are based upon the theory of statistics and machine learning learning techniques. Anomalies, patterns and trends are modelled from historical data collected from a monitored system, then the trends are utilised to estimate RUL of the system.

As described in section 1.2.3.1, a gas turbine engine is a complex system where it is difficult to be modelled by physics-based as well as knowledge-based approaches. On the other hand, a considerable amount of data is available because modern engines have been fitted by several EHM sensors and there is a large number of engines to be monitored. Hence, data-driven approaches will be appropriate methods for gas turbine engine prognostics.

According to Goebel et al. [2008a]; Heng et al. [2009]; Schwabacher and Goebel [2007]; Sikorska et al. [2011], one of the most popular methods for data-driven prognostics is artificial neural networks. An artificial neural network is a computational network which attempt to simulate the networks of nerve cell (neurons) of the biological (human or animal) central nervous system [Graupe, 2007]. This approach has been used for remaining life predictions in many applications, such as electro-hydraulic servo

valve of aircraft actuator components [Byington et al., 2004a], planetary gear plate of a helicopter transmission [Khawaja et al., 2005], planetary gear train of motor-pump in power station [Yam et al., 2001] and bearing [Gebraeel and Lawley, 2008; Huang et al., 2007; Shao and Nezu, 2000]. However, in the field of safety-related applications it is essential to provide transparent solutions that can be validated by domain experts where neural network is considered as a "black box" approach [Nusser, 2009] and ultimately it is difficult to determine the most appropriate model. The biggest issue with using most of neural networks[1] for RUL prediction is that they do not naturally provide confidence limits for their output predictions [Sikorska et al., 2011].

Other examples of data-driven approaches include linear regression [Li and Nilkitsaranont, 2009], Dempster-Shafer theory [Goebel et al., 2006], autoregressive models [Wu et al., 2007; Yan et al., 2004], regression tree [Tran et al., 2008], Hidden Markov Model [Zhang et al., 2005], Support Vector Machine [Caesarendra et al., 2010] and Bayesian approaches [Gebraeel et al., 2005; Przytula and Choi, 2008].

Data-driven approaches have been developed for gas turbine engine prognostics. For example, Marinai [2004]; Marinai et al. [2003] applied two techniques to handle different problems in RUL estimation of gas turbine engine. The Box-Jenkins autoregressive integrated moving average (ARIMA) method was implemented to provide accurate forecasts for immediate and short-term forecasting, whereas regression analysis was used to handle prognostics that require medium- and long-term predictions. However, their proposed approaches require numerous data from a single individual unit to produce realistic prediction at short horizon[2]. Prior to it, the underlying models do not have the capability to hold the belief about engine degradation. These methods are also not based on probabilistic framework, leading to the absence of uncertainty around the prediction.

Each approach has its own advantages and disadvantages as discussed in the above survey papers. This requires mathematical understanding of each model type, and also an appreciation of how a particular business intends to utilise the models and their outputs [Sikorska et al., 2011].

The proposed methodology is to select the most suitable prognostic approaches for

[1]Here, it means most neural network applied to prognostics, where they are not based on probabilistic framework.

[2]This approach may be inefficient as it requires large amount of historical data from a single unit to predict a short horizon.

complex system's problem by reviewing data-driven techniques from the mature domains related to prognostics. A wide viewpoint has been adopted whilst reviewing the literature related to prognostics. This macroscopic view has uncovered three different classes of technique related to data-driven prognostics: reliability engineering, survival analysis and forecasting. These fields have been explored in the context of the disciplines of healthcare and finance, as well as in engineering.

2.3.3.1 Reliability engineering

Reliability is the ability of a system or component to perform its required functions under stated conditions for a specified period of time [Hamada et al., 2008]. Reliability theory has been widely used in the aviation industry to estimate the failure time probability distributions for aircraft fleets. Examples of reliability theory include: the use of the Bernstein reliability model to model the life characteristics of machine components by Ahmad and Sheikh [1984], and the development of statistical methods for using degradation measures to estimate a time-to-failure distribution, described in Lu and Meeker [1993]. The reliability failure distribution can be used as initial (prior) probability distributions (the knowledge base) in the prognostic process to reduce uncertainty. This field is concerned with inferences based upon a population of assets.

2.3.3.2 Survival analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is the time until an event occurs [Kleinbaum and Klein, 2005]. In contrast to the population approach prevalent in reliability theory, the techniques described as survival analysis are often more individual-centred, especially in biomedical applications. For instance, research into the survival of multiple myeloma patients which examined the association between the values of certain explanatory variables or covariates and the survival time of patients is conducted by Medical Centre of the University of West Virginia, USA [Collett, 2003]. Another field related to survival analysis is named competing risks analysis. Competing risks analysis is a field of applied statistics that has the ability to handle dependent failure. The experts in this area claim that "if something can fail, it can often fail in one of several ways and in more than one way at a time" [Ma and Krings, 2008]. The unique capability of com-

peting risks analysis in handling information censoring and dependent failures makes this approach potentially valuable for analysing and modelling time-to-event data. In analysing survival data, information censoring, such as the incomplete observation of survival times, is a particular challenge. Survival analysis has already tackled some of these problems and there is a portfolio of mathematical models and methodologies that have been developed to extract information from the censored observations maximally [Ma, 2008].

2.3.3.3 Forecasting

Forecasting is the construction of a suitable model based upon analysis of the historical development of a data series and utilisation of information relevant to its likely future development [Pole et al., 1994]. This discipline is very common in econometrics, such as commodity price modelling and consumption predictions. In addition, other fields of science, e.g. meteorology, have obtained benefit from this theory [Makridakis et al., 2008]. Evidence of numerous forecasting methods have been applied to prognostics, for example, the use of ARIMA models to predict future machine health [Wu et al., 2007] and prognostics of remaining bearing life using neural networks [Shao and Nezu, 2000]. Forecasting methods, through the use of sensor data, are adept at capturing the idiosyncratic failure behaviour of real world components.

2.3.3.4 Bayesian framework

It is believed that the advantages of the aforementioned fields should be taken into consideration to develop optimal prognostic methods because each discipline has unique capabilities and past successes in research. Figure 2.4 shows the interaction of prognostics with these fields. Prognostics can be considered as an intersection of the combination of these disciplines and therefore this representation is a powerful illustration of approaches that might be applied in prognostics.

Bayesian approaches have been utilised widely and successfully in many areas related to prognostics. The examples include reliability engineering [Hamada et al., 2008], survival analysis [Ibrahim et al., 2001] and forecasting [Pole et al., 1994; West and Harrison, 1997]. From those perspectives, Bayesian methods can be considered as an intersection of multidisciplinary fields, which are reliability, survival analysis and fore-

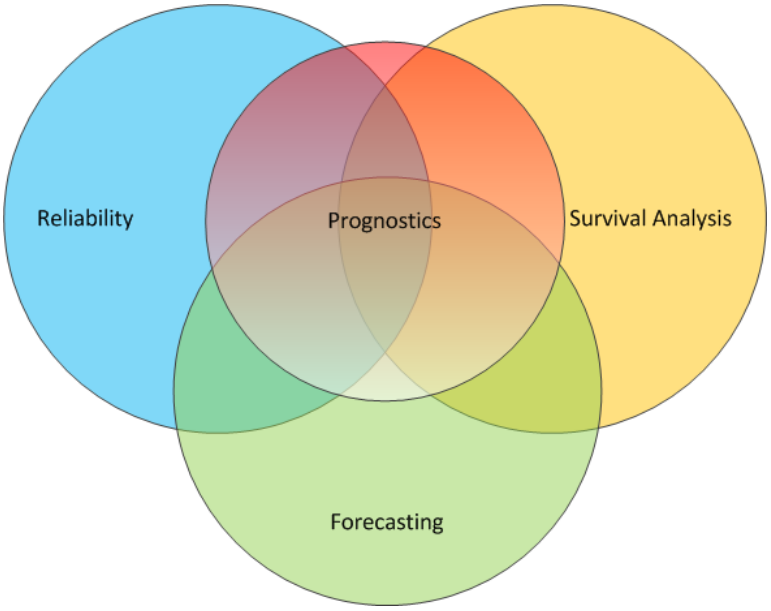


Figure 2.4: The interaction of prognostics with different disciplines. The use of a particular algorithm which has been used successfully in these fields is beneficial to achieve prognostic goal.

casting, as shown in figure 2.4. Hence, our research is devoted on developing prognostic algorithms within Bayesian framework.

In the area of prognostics, several researchers have adopted machine learning techniques in Bayesian framework, such as Relevance Vector Machine (RVM) and Bayesian network. For example, Przytula and Choi [2008] proposed a probabilistic approach in the form of dynamic Bayesian network for reasoning in diagnostics and prognostics. The approach had been tested on several examples of health prognosis for electromechanical and electronic subsystems in aviation. In addition, Widodo and Yang [2011a] used RVM to predict survival probability of individual bearings. RVM was trained from input data obtained from run-to-failure data[1] with survival probability estimated by Kaplan-Meier (KM) and probability density function estimators as target. However, both approaches used run-to-failure data for training the algorithms, which is impractical for our case, because gas turbine engine has relatively long lifetime and it is never allowed to fail. Instead, we utilise degradation data, that is TGT margin for the RUL estimation (section 1.2.2).

[1]Run-to-failure data is times to the occurrence of a certain event, such as fault and failure.

Furthermore, Saha and Goebel [2008]; Saha et al. [2007] utilised Bayesian techniques in the form of RVM combined with a Particle Filter to provide diagnostics and prognostics of battery health. However, their research focus was an electronic system, for which they used a physics-based model. Therefore, they must develop realistic physics of failure model by deriving from lumped parameter model of a battery cell. The idea of this research is not really applicable for complex system applications which are difficult to model based on physical equations.

In gas turbine engine prognostics, Lipowsky et al. [2010] applied Bayesian forecasting and Dynamic Linear Models (DLMs) for capturing uncertainty in degradation data. This method detects the change in the data (outlier detection) and perform a prognostics of measurement values. The drawbacks of this approach are that several crucial parameters need to be determined heuristically, leading to uninformative distributions and while detecting the change, "rule of thumb" is involved to determine gradient change in the prognostic trajectory. The use of "rule of thumb" leads to non-generic approach, i.e. the method may only work for a very specific problem.

Gebraeel et al. [2005] and Gebraeel [2006] developed a Bayesian method to update on-line the stochastic parameter of exponential degradation models for bearing application. This method is very promising because it utilises available reliability data to form an informative prior, i.e. the initial knowledge about degradation. This results in an appropriate RUL estimation in early prediction. The developed Bayesian method is also able to update failure time predictions by the appropriate use of real-time condition monitoring information and reliability information.

Gebraeel and Pan [2008] then extended the research to model degradation mathematically in a time-varying environment. Chakraborty et al. [2009] continued the same research by involving non-symmetric priors, such as the Gamma prior in a Bayesian framework. This can be applied for a special case when the underlying normality assumptions are not satisfied. For instance, when the prior distribution of the stochastic parameter is skewed, therefore Gamma distribution is selected due to the flexibility in capturing the characteristics of the real-world sensory data. In complementary work, Elwany and Gebraeel [2009] modelled the degradation signal using a random coefficient model with error terms that follows a Brownian motion process. Then, they approximated the procedure of evaluating conservative mean of the sensory-update Remaining Life Distributions (RLDs) and expressed the mean and variance using closed-form ex-

pressions that are easy to evaluate.

The prognostic approaches, developed by Gebraeel [2006]; Gebraeel et al. [2005], are more generic rather than their following methods. Therefore, this research had been initiated by adopting their basic concept. The next section will deliver a general description of Bayesian regression technique.

2.4 Prognostics using Bayesian methods

In statistics, Bayesian inference is a method of inference in which Bayes' rule is used to update the probability estimate for a hypothesis as additional evidence is acquired. Bayes' theorem is named after Thomas Bayes (1701 - 1761), who first suggested using the theorem to update beliefs. In fact, Bayes only formulated his theory for the case of a uniform prior, and it was Pierre-Simon Laplace (1749 - 1827) who independently rediscovered the theory in general form and who demonstrated its broad applicability [Bishop, 2006].

Bayesian methods are able to improve the quality of the models by incorporating a *priori* qualitative and quantitative knowledge. In other words, Bayesian approaches allow us to assign the prior distributions to the parameters in the model which capture known qualitative and quantitative features, and then to update these priors in the data, yielding a posterior distribution via Bayes' theorem [Denison et al., 2002]:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior} \tag{2.1}$$

2.4.1 Motivation

Three main factors have driven the focus on a Bayesian framework for data-driven prognostic approaches:

- Bayesian method provides estimates of probability distributions, which have advantages over point value estimates, such as quantifying the risk of failure, handling uncertainty, etc [Bishop, 2006]. For example, prognostics using a neural network[1] [Gebraeel et al., 2004], estimates RUL as a single time point while, by

[1] Here, the used neural network was not based on probabilistic framework.

using a Bayesian method, RUL estimation takes the form of a failure-time distribution. As a result, any decision can be informed by the entire distribution; hence the risk of early failure can be assessed.

- Bayesian approach is transparent in their assumptions, where it gives a transparent view of the rules that govern the relationships that make predictions possible [Marwala and Crossingham, 2008; O'Hagan and Luce, 2003]. It is important to provide transparent solution which can be validated by experts in the field of safety-related applications [Nusser, 2009].

- Bayesian framework allows for the incorporation of prior distribution (previous experience/knowledge) in a coherent way and avoids over-fitting problems [Beal, 2003]. This approach differs from conventional methods, such as least-squares estimation, because its ability to update the probability of future observations by incorporating evidence from previous experience and experiments into the overall conclusion [Berry, 1996]. This combined information can be used to estimate model coefficients that result in generalised degradation models. Therefore, Bayesian technique is able to cope with large uncertainty problem (section 1.2.3.3), which result in appropriate predictive distributions to estimate RUL, especially in early prediction, when there is little or no data, available.

2.4.2 Bayesian Regression

In statistics, Bayesian linear regression is a statistical method for linear regression within the context of Bayesian inference. In this section, the Bayesian treatment of the linear model, which has also been used by Gebraeel [2006]; Gebraeel et al. [2005] for bearing prognostics, is discussed.

2.4.2.1 Model

A standard linear regression model with Gaussian noise is defined by:

$$\begin{aligned} y &= \sum_{d=0}^{D-1} \phi_d(\mathbf{x}) w_d + \varepsilon \\ &= \boldsymbol{\phi}(\mathbf{x})\, \mathbf{w} + \varepsilon \end{aligned} \tag{2.2}$$

where D is the total number of parameters in this model, $\phi_d(\mathbf{x})$ are known as basis functions, y is output variable and $\mathbf{x}$ is input variables. ε is a random error term that follows a Gaussian distribution[1], $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ and σ^2 is noise variance. $\boldsymbol{\phi}(\mathbf{x}) = [\phi_0(\mathbf{x}),\, \phi_1(\mathbf{x}), \cdots, \phi_{D-1}(\mathbf{x})]$ and $\mathbf{w} = [w_0,\, w_1,\, \cdots, w_{D-1}]^T$. The parameter w_0 allows for any fixed offset in the data[2], so that an additional dummy "basis function", $\phi_0(\mathbf{x}) = 1$, is defined [Bishop, 2006].

If there are N data points, applied to the above model, the model can be written as:

$$y_i = \boldsymbol{\phi}(\mathbf{x}_i)\, \mathbf{w} + \varepsilon_i \tag{2.3}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi}(\mathbf{x}_1) \\ \boldsymbol{\phi}(\mathbf{x}_2) \\ \vdots \\ \boldsymbol{\phi}(\mathbf{x}_N) \end{pmatrix} \mathbf{w} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix} \tag{2.4}$$

where $i = 1,\, 2,\, \cdots, N$. This model can be simplified as:

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{w} + \boldsymbol{\varepsilon} \tag{2.5}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\mathbf{x}_1),\, \boldsymbol{\phi}(\mathbf{x}_2), \cdots, \boldsymbol{\phi}(\mathbf{x}_N)]^T$, which is a design matrix. $\boldsymbol{\Phi} \triangleq \boldsymbol{\Phi}(X)$, where $\mathbf{X} = [\mathbf{x}_1,\, \mathbf{x}_2, \cdots, \mathbf{x}_N]^T$ is a data set of inputs with corresponding output values, $\mathbf{y} = [y_1,\, y_2, \cdots, y_N]^T$ and $\boldsymbol{\varepsilon} = [\varepsilon_1,\, \varepsilon_2, \cdots \varepsilon_N]^T$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I)$.

2.4.2.2 Prior distribution

A key difference between Bayesian inference and the maximum likelihood estimate lies in prior distribution. Bayesian regression uses prior distribution for the inference. In this case, a Gaussian distribution is used as a prior distribution, given by:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\bar{\mathbf{w}}, V) \tag{2.6}$$

[1] A Gaussian distribution is also known as a Normal distribution (appendix A.2).
[2] This is sometimes called a bias parameter.

where $\bar{\mathbf{w}}$ and V are mean and covariance of $\mathbf{w}$, respectively. Gaussian distribution is chosen as prior distribution because it is a conjugate prior[1] with respect to a Gaussian likelihood function [Bishop, 2006].

2.4.2.3 Likelihood function

The likelihood for this model is conditional on the joint probability of observing the data ($\mathbf{X}$) and the model parameter ($\mathbf{w}$). This can be written:

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \mathcal{N}(\mathbf{y}|\mathbf{\Phi}\mathbf{w}, \sigma^2 I) \tag{2.7}$$

2.4.2.4 Posterior distribution

Having the likelihood function and the prior distribution, the posterior distribution can be computed using Bayes' theorem:

$$\underbrace{p(\mathbf{w}|\mathbf{y}, \mathbf{X})}_{\text{posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{X}, \mathbf{w})}_{\text{likelihood}} \underbrace{p(\mathbf{w})}_{\text{prior}} \tag{2.8}$$

Bayes's theorem for Gaussian variables stated in appendix A.2 can be used, resulting in:

$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) = \mathcal{N}(\mathbf{w}|\mathbf{w}^*, V^*) \tag{2.9}$$

where $\mathbf{w}^*$ and V^* are mean and covariance of posterior $\mathbf{w}$, respectively.

$$\mathbf{w}^* = V^* \left(V^{-1}\bar{\mathbf{w}} + \mathbf{\Phi}^T\mathbf{y}/\sigma^2\right) \tag{2.10}$$

$$V^* = \left(V^{-1} + \mathbf{\Phi}^T\mathbf{\Phi}/\sigma^2\right)^{-1} \tag{2.11}$$

2.4.2.5 Predictive distribution

Predictive distribution is predictions of $\mathbf{y}^*$ for new values of $\mathbf{X}^*$, which can be obtained by marginalising the posterior distribution:

$$p(\mathbf{y}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \int p(\mathbf{y}^*|\mathbf{X}^*, \mathbf{w})\, p(\mathbf{w}|\mathbf{y}, \mathbf{X})\, d\mathbf{w} \tag{2.12}$$

[1] The use of a conjugate prior results in a posterior distribution having the same functional form as the prior, and therefore lead to a greatly simplified Bayesian analysis.

This involves the convolution of two Gaussian distributions, and so making use of the result in appendix A.2, gives the predictive Gaussian distribution:

$$p(\mathbf{y}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\mathbf{y}^*|\mathbf{\Phi}^*\mathbf{w}^*, \mathbf{\Phi}^* V^* \mathbf{\Phi}^{*T} + \sigma^2 I) \tag{2.13}$$

2.4.2.6 Advantages

This type of Bayesian regression has an analytical solution because it involves conjugate Gaussian distributions. In prognostics, this approach has two main advantages:

- Based on available data, parameters in a Gaussian prior are relatively easy to determine, resulting in an informative prior distribution.
- Solution for posterior and predictive distributions can be computed deterministically.

2.5 Summary

This chapter presents a literature review and discusses briefly various aspects related to prognostics. Firstly, the explanation of the factors influencing prognostics accuracy is provided to understand the process of PHM in general. The prognostic process represents a complete prognostic system, required to be deployed in field applications. One of the most influential accuracy factors is the prognostic algorithms, which is the main focus of the research.

Subsequently, the literature review about prognostic algorithms is described based on our classification: physics-based, knowledge-based and data-driven approaches. The advantages and drawbacks of each approach are discussed. An overview highlighting similarities between several disciplines is also presented and common connections to prognostics are identified. The interaction diagram is helpful to classify prior art from different domains and help us to select an appropriate approach which has been used widely and successfully in those fields.

Bayesian methods have been identified to possess many appropriate traits. Therefore, we describe briefly a standard Bayesian regression method in the final section. This method is a promising method which will be the benchmark for developing prognostic algorithms in subsequent chapters.

In the next chapter, the description will be about: how this Bayesian technique can be implemented based on prognostic methodology described in Gebraeel [2006]; Gebraeel et al. [2005]. Subsequently, this algorithm is compared with the developed algorithm using case studies.

Chapter 3

Bayesian Regression: a deterministic approach

3.1 Introduction

Bayesian approaches are potentially very effective means to perform prognostics, especially when a physically-based degradation model does not exist. However, the existing Bayesian Regression (BR) method [Gebraeel, 2006; Gebraeel et al., 2005], which is also described in section 2.4.2, assumes that the variation in observable health index is constant over the life of the asset, i.e. the estimation errors are assumed to have a fixed noise variance. This leads to inflexible uncertainty in capturing noise characteristics.

To overcome this limitation, the method proposed in this chapter extends the aforementioned Bayesian method by modelling the noise variance as a random variable[1]. Their method can be called as Bayesian Regression 1 (BR-1) whereas the proposed method is named Bayesian Regression 2 (BR-2).

3.2 Bayesian Regression 2

This section describes the use of the Bayesian framework for prognostic analysis, where the following discussion explains specifically the BR-2 approach.

[1] A random variable is a variable which takes values based on a certain probability distribution [Amemiya, 1994].

3.2.1 Model and block diagram

BR-2 uses the same model as BR-1. Recall BR-1 model, equation (2.5), the degradation signal (the health index), $\mathbf{y}$, can be modelled as:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\varepsilon} \tag{3.1}$$

where $\boldsymbol{\varepsilon}$ is a random error term that follows a Gaussian distribution, $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I)$ and σ^2 is noise variance. $\mathbf{\Phi}$ is $N \times D$ design matrix of basis function, which maps an N dimensional input vector into a D dimensional feature space and $\mathbf{w}$ is a D dimensional vector of weights.

Similarly to BR-1, this technique also combines two sources of information, the reliability of the component population and the real-time sensor information to update periodically the RUL of a unit. The concept can be calculated using Bayes' theorem, equation (2.1), that is: posterior $\propto$ likelihood $\times$ prior.

There are five main components in Bayesian prognostics. Figure 3.1 illustrates the diagram of updated Bayesian prognostic methods for BR-2 and BR-1. The first three components are used to process the health signal. A degradation database, which contains a collection of degradation data, is used to estimate parameters in prior distribution. Observed degradation data are modelled to be a likelihood function. Having prior distribution and likelihood function, posterior distribution can be computed. Subsequently, the predictive distribution can be calculated and this is then predicted over the desired time horizon. In the final step, this is transformed into a failure-time distribution to deliver an estimate of RUL. The foundations of BR-2 have been developed from the ideas expressed in Denison et al. [2002]; Gelman et al. [2004]; Rossi et al. [2005].

The choice of prior distribution is a key difference between the BR-2 and BR-1 approaches. Unlike BR-1 which assumes the noise variance as a fixed parameter, BR-2 treats the noise variance as a random variable because the variation of health index may change either in time or across units.

3.2.2 Prior distribution

It is necessary to select an appropriate prior distribution in order to obtain an analytically tractable posterior distribution, which is desirable for real-time, deterministic computa-

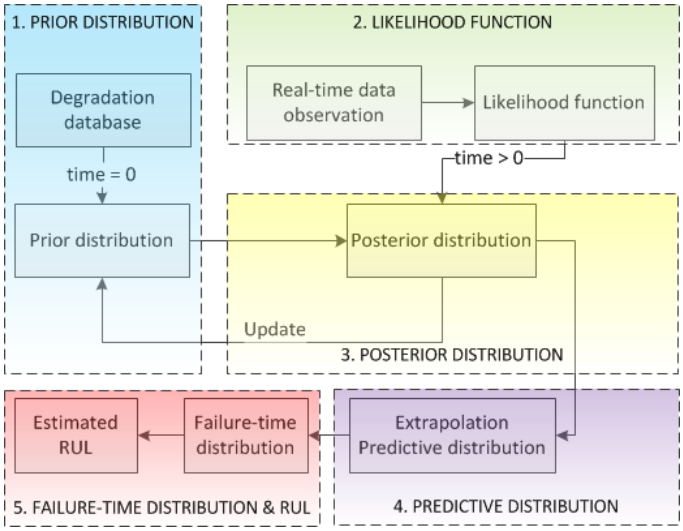


Figure 3.1: Diagram of the updated Bayesian method for BR-1 and BR-2. Parameters in prior distribution are estimated from degradation database and the observed degradation data are modelled to be a likelihood function. Thus, posterior distribution can be calculated by Bayes' rule. Subsequently, predictive distribution is computed and extrapolated until it crosses the defined threshold. Finally, failure-time distribution is obtained and therefore RUL can be estimated.

tion. BR-2 uses two types of probability distribution to model both the weights and the noise variance. The weights are modelled as a Gaussian distribution to capture variable dependency, whilst the noise variance is assumed as a random variable following an Inverse Gamma (IG) distribution which provides more flexibility than BR-1 in capturing the noise characteristic. Hence, the prior distribution is specified as:

$$p(\mathbf{w}, \sigma^2) = p(\mathbf{w}|\sigma^2)p(\sigma^2) \tag{3.2}$$

The component $p(\mathbf{w}|\sigma^2)$ is prior of $\mathbf{w}$, which as before is chosen to be the conjugate, Gaussian distribution. The second term $p(\sigma^2)$ is the marginal prior of σ^2, which has a density on the form of IG distribution. The outcome is a natural conjugate prior $p(\mathbf{w}, \sigma^2)$ which is a Normal Inverse Gamma (NIG) distribution. Thus, this prior can be written as:

$$p(\mathbf{w}, \sigma^2) = \text{NIG}(\bar{\mathbf{w}}, V, a, b) \tag{3.3}$$

where $\bar{\mathbf{w}}$ and V are mean and covariance of the weight prior. Parameters a and b are the parameters of the IG distribution.

In general, prior parameters can be estimated from non-informative prior (diffuse prior) or available data (informative prior) or expert knowledge (subjective prior) [Kass and Wasserman, 1996]. In this subsection, we propose a methodology where prior parameters for BR-2 can be estimated fully from in-service database (highly informative prior).

In BR-2, there are two prior distributions with four parameters to be specified. First prior is Gaussian prior distribution, $p(\mathbf{w}|\sigma^{\mathbf{2}})$. The parameters $\bar{\mathbf{w}}$ and V can be computed by taking the mean and covariance of all of $\mathbf{w}$ estimated, e.g. by Ordinary Least Square (OLS).

Another prior is IG distribution, $p(\sigma^2)$, where a method to produce a highly informative IG prior distribution is introduced. To estimate its prior parameters, a and b, the squared residual value between predicted signal using OLS and the observed degradation is calculated for each unit. Then, the mean ($\hat{\mu}$) and variance ($\hat{\sigma}^2$) of the squared residual value can be computed. The parameters a and b can then be calculated by:

$$a = \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 + 2 \tag{3.4}$$

$$b = \hat{\mu}\left(\left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 + 1\right) \tag{3.5}$$

The proposed methodology for specifying parameters in IG prior distribution will have a big impact in a "more" complex Bayesian model as we shall see in chapter 4.

3.2.3 Likelihood function

The likelihood for a model is the conditional probability of the observed data ($\mathbf{X}$) and the model parameters ($\mathbf{w}, \sigma^2$). Hence, this can be written:

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) = \mathcal{N}(\mathbf{\Phi w}, \sigma^2 I) \tag{3.6}$$

3.2.4 Posterior distribution

Having the likelihood function and the prior distribution, the posterior distribution can be computed using Bayes' theorem. It can be expressed as:

$$p(\mathbf{w}, \sigma^2|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2)p(\mathbf{w}|\sigma^2)p(\sigma^2) \tag{3.7}$$

Owing to the conjugate choice, the posterior distribution is also obtained in the same form as the prior: a NIG distribution.

$$p(\mathbf{w}, \sigma^2|\mathbf{y}, \mathbf{X}) = \text{NIG}(\mathbf{w}^*, V^*, a^*, b^*) \tag{3.8}$$

where $\mathbf{w}^*$, V^*, a^* and b^* represent the updated posterior parameters for the NIG distribution. The posterior parameters can be updated as more data ($\mathbf{X}$) emerges through the following equations [Denison et al., 2002; Murphy, 2007]:

$$\mathbf{w}^* = \left(V^{-1} + \mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\left(V^{-1}\bar{\mathbf{w}} + \mathbf{\Phi}^T\mathbf{y}\right) \tag{3.9}$$

$$V^* = \left(V^{-1} + \mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1} \tag{3.10}$$

$$b^* = b + \frac{1}{2}\left(\bar{\mathbf{w}}^T V^{-1}\bar{\mathbf{w}} + \mathbf{y}^T\mathbf{y} - (\mathbf{w}^*)^T(V^*)^{-1}\mathbf{w}^*\right) \tag{3.11}$$

$$a^* = \frac{n}{2} + a \tag{3.12}$$

where the derivation of posterior distribution for BR-2 is discussed in detail in appendix B.1. The resulting posterior distribution is then used to update the prior for the next iteration and the posterior is also utilised to compute the predictive distribution.

3.2.5 Predictive distribution

Once the posterior distributions have been updated through equations (3.9)-(3.12), the goal is to determine the predictive distribution (component 4 of figure 3.1). This distribution represents the predicted degradation signal, $\mathbf{y}^*$, probabilistically. BR-2 results in a predictive distribution that takes the form of a Student-t distribution [Bishop, 2006; Denison et al., 2002] and this is given by:

$$p(\mathbf{y}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) \sim \text{St}(\mathbf{\Phi}^*\mathbf{w}^*, b^*(I + \mathbf{\Phi}^* V^* \mathbf{\Phi}^{*T}), a^*) \tag{3.13}$$

where $\mathbf{X}^*$ is a new data set of inputs. The derivation of predictive distribution is discussed in appendix B.2.

In the current problem, this will typically correspond to an extrapolation of the degradation signal $\mathbf{y}^*$. The predictive distribution is extrapolated until a relevant statistic (e.g. the mean) crosses the defined failure threshold. This can be then converted into a failure-time distribution.

3.2.6 Failure-time distribution and RUL

The failure-time distribution $f(t)$ can be represented by a Bernstein distribution [Ahmad and Sheikh, 1984; Gebraeel et al., 2005]. The probability that the estimated failure time (T) exceeds time (t) is equivalent to the probability that the future degradation signal (y^*) is less than some failure threshold (y_{fail}). This can be formulated as:

$$P(T > t) = P(\mathrm{y}^* \leq \mathrm{y}_{\text{fail}}) \tag{3.14}$$

Therefore, the failure-time distribution can be written as:

$$\begin{aligned} F(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P(\mathrm{y}^* \leq \mathrm{y}_{\text{fail}}) \\ &= 1 - F_a\left(\frac{\mathrm{y}_{\text{fail}} - \mathbf{\Phi}\mathbf{w}^*}{\sqrt{b^*(I + \mathbf{\Phi}V^*\mathbf{\Phi}^T)}}, a^*\right) \end{aligned} \tag{3.15}$$

$$f(t) = \frac{dF(t)}{dt} \tag{3.16}$$

where $F_a(.)$ is the cumulative distribution function (cdf) of the Student-t distribution [Shaw, 2006]. Because $f(t)$ is derived from Student-t distribution, it can be regarded as an extension of a Bernstein distribution.

As defined in Saxena et al. [2008a], the RUL is the amount of time remaining before the degradation signal crosses a defined failure threshold. Therefore, the final step requires estimation of RUL by subtracting the current time and the mode of failure-time distribution $f(t)$. The time at the mode of $f(t)$ is called the estimated failure time, denoted by T, and the current time is denoted by τ. The RUL is then given by: $\text{RUL} = T - \tau$.

3.2.7 Algorithm summary

In summary, figure 3.2 shows the graphical models for BR-1 and BR-2. The main difference between these models can be seen clearly in the treatment of the observation noise variance σ^2. In BR-2, σ^2 is modelled as a random variable following an IG distribution, while in BR-1, σ^2 is treated as a fixed parameter.

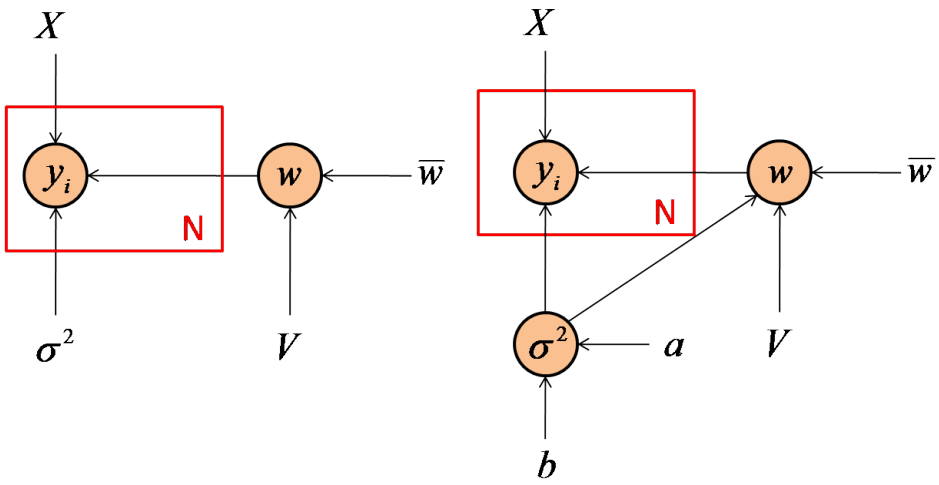


Figure 3.2: Graphical models illustrating the differences between BR-1 (left) and BR-2 (right). BR-2 treats the noise variance, σ^2, as a random variable, while in BR-1, σ^2 is treated as a fixed parameter.

Figure 3.3 shows a block diagram for a whole methodology to run the proposed prognostic algorithm. Once prior parameters have been estimated for all units, each unit is fed to a separated BR-2 algorithm, where the details of BR-2 block are shown in figure 3.1. The BR-2 method can be updated as each new observed datum arrives. This process continues until a decision to replace or maintain is made (when the observation is completed). For each degradation cycle[1] which has been observed fully[2], this data will be added to degradation database. Therefore, when a new unit comes, its prior parameters can be estimated from the updated degradation database.

[1]A degradation cycle contains complete observed degradation data for one unit. For a recoverable system, this includes complete degradation data between two maintenances.

[2]Here, there are two possibilities why we stop data observation (collection). First possibility is when the system fails whereas another possibility is when there is a maintenance activity (for a recoverable system).

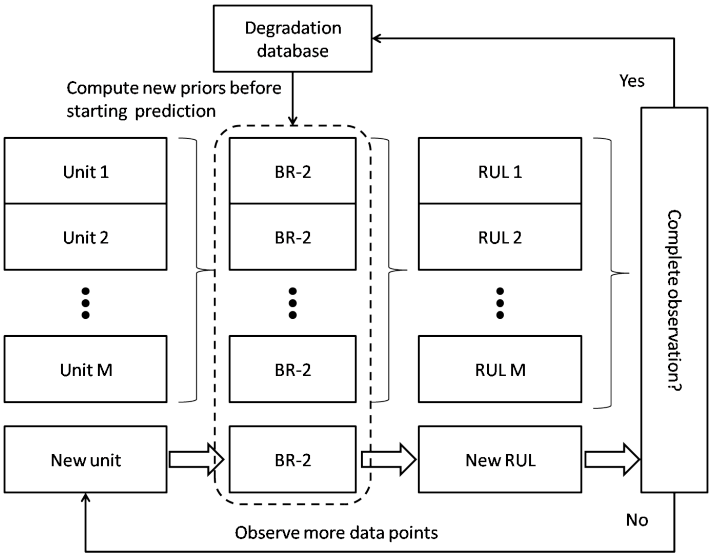


Figure 3.3: A complete block diagram of updated Bayesian method. Once the prior parameters have been estimated for all units from degradation database, RUL of multiple units are estimated independently by BR-2 algorithm. Degradation database is updated by adding the complete observation of a degradation cycle. If a new unit is introduced, its prior parameters can be estimated from an updated degradation database.

3.3 Case Studies

This section describes a comparison of the two prognostic algorithms: BR-1 and BR-2. In this case study, prognostics focuses on a degradation cycle. Therefore, degradation data between two maintenance events are selected and any maintenance effect on prognostics are neglected. The effect of maintenance events into prognostics will be discussed in chapter 6.

As described in section 1.2.3.2, real TGT margin degradation data is difficult to use, directly, for comparing prognostic algorithms, because they are usually right-censored and so do not have ground truth for failure time. In the first case study, therefore, synthetic degradation data of known properties is generated to test the algorithms. Subsequently, in the second case study, the algorithms are applied to the real TGT margin data for validation.

3.3.1 Case study 1: Synthetic degradation data

Here the algorithms are tested on synthetic data which mimics noisy degradation data. This data can be used to create many scenarios for prognostics since the model is known and its parameters can be controlled. The synthetic data provides a powerful method to validate the performance of a given algorithm prior to testing on a real prognostic case.

The synthetic data is generated to model real data characteristics using two linear models with various levels of noise, given by:

$$\mathbf{d}_1 = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\varepsilon}_1 \tag{3.17}$$

$$\mathbf{d}_2 = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\varepsilon}_2 \tag{3.18}$$

where $\mathbf{w}$, $\boldsymbol{\varepsilon_1}$ and $\boldsymbol{\varepsilon_2}$ are each drawn from a Gaussian distribution, so that: $\mathbf{w} \sim \mathcal{N}(\mu_{\mathbf{w}}, \Sigma_{\mathbf{w}})$, $\boldsymbol{\varepsilon_1} \sim \mathcal{N}(0, \sigma_1^2 I)$ and $\boldsymbol{\varepsilon_2} \sim \mathcal{N}(0, \sigma_2^2 I)$. Here, first order polynomial basis function[1] is used, $\phi_d = x^d$, so that $\mathbf{\Phi} = [\mathbf{1}\ \mathbf{x}]$, where $\mathbf{1}$ and $\mathbf{x}$ are N dimensional column vector of $1s$ and input variable, respectively. The mean of $\mathbf{w}$ can be decomposed as $\mu_{\mathbf{w}} = [\mu_{\mathbf{w}_1}\ \mu_{\mathbf{w}_2}]^T$. It is necessary to constrain $\mu_{\mathbf{w}_2} < 0$, to ensure that the synthetic data degrade downward (having negative slope)[2].

This case study attempts to produce two main scenarios in order to investigate the key differences between BR-1 and BR-2, for prognostic problem. Ten degradation signals are generated using each model. Units 1-10 are generated by the first model, $\mathbf{d}_1$, where the noise variance, σ_1^2, is kept the same for all units, whereas units 11-20 are generated using the second model, $\mathbf{d}_2$, where noise variance is varied, σ_2^2, from one unit to another unit, so that $\boldsymbol{\varepsilon_{2j}} \sim \mathcal{N}(0, \sigma_{2j}^2)$, where j indexes unit j. Let assume a set of noise variances for the second model as $\boldsymbol{\sigma}_2^2 = [\sigma_{21}^2, \cdots, \sigma_{2M}^2]$, where M is the total number of unit. In this case, noise variance on the first model is selected as $\sigma_1^2\, \epsilon\, \boldsymbol{\sigma}_2^2$, meaning that, σ_1^2 is a member of the set $\boldsymbol{\sigma}_2^2$.

Figure 3.4 shows 20 synthetic degradation signals, generated by two degradation models. Time index is used to represent the (normalised) time, and the synthetic health index embodies degradation signal (normalised). It can be seen that unit 1 is symbolised

[1]This assumption follows the explanation in section 1.2.3.5, that degradation patterns in gas turbine engine are nearly linear.

[2]This assumption mimics TGT margin degradation as will be shown in real case study (see also figure 1.5).

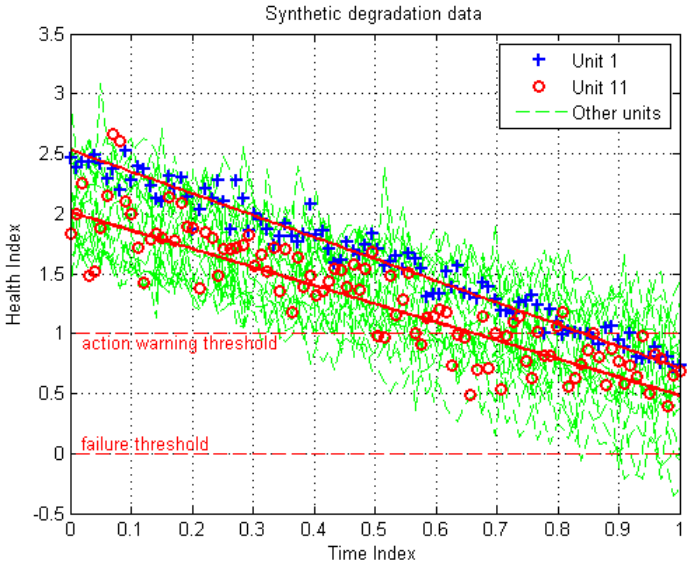
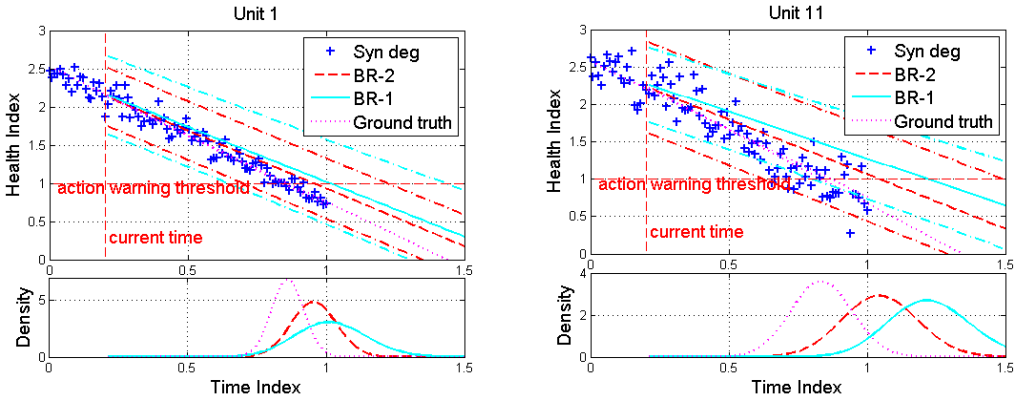


Figure 3.4: Synthetic degradation data with low and high levels of noise for 20 units. The symbols "+" (unit 1) and "○" (unit 11) represent degradation signals for low and high level noise, respectively. The other degradation signals are shown by dashed lines.

by "+", unit 11 is symbolised by "○" whereas their ground truths are shown by solid lines. The other degradation signals are shown by dashed lines. In this case, the action warning threshold is assumed to be one, whereas the failure threshold is assumed to be zero. A failure might occur when health index crosses the failure threshold. However, in reality the health index should never be allowed to cross the failure threshold. Therefore, RUL is calculated based on the action warning threshold, i.e. the estimates predict when an action warning will be issued.

In this case study, there are two different scenarios of prognostics. In the first scenario, prognostics is applied to units 1-10, where their noise variances are fixed whilst in another scenario, prognostics is implemented to units 11-20, in which the noise variances vary between units and the noise level may be lower or higher than the first scenario.

For each scenario, nine degradation signals are used for training the prior distribution, $p(\mathbf{w}, \sigma^2)$, using the OLS method, as described in section 3.2.2. Subsequently, both prognostic algorithms are used to estimate RUL on the remaining unit. Here, first-order polynomial basis function, $\phi_d = x^d$, is also used, where it can be written as a matrix



(a) (b)

Figure 3.5: Representation of trajectory prediction for unit 1 (left) and unit 11 (right), when few data are available at time index 0.2. For unit 1, both predictions show "similar" performance, because in this case, all units share the same noise variance with low level of noise. BR-1 prediction for unit 11 shows a slight deviation from the true value as it contains high level of noise and the noise also varies across units, whilst BR-2 prediction is able to retain the prediction close to the truth.

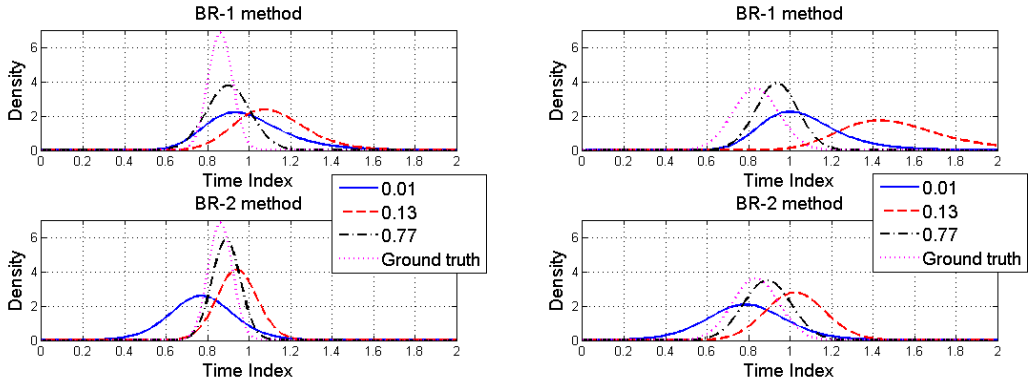
design, given by:

$$\Phi = [\mathbf{1}\ \mathbf{x}] \tag{3.19}$$

where $\mathbf{1}$ and $\mathbf{x}$ are N dimensional column vector of $1s$ and input variable, respectively.

Figure 3.5 demonstrates the trajectory (degradation) predictions for the synthetic data for units 1 and 11 at time index 0.2. In the top subfigures, the symbol, "+", is the health index value corrupted by the noise whereas the straight dotted line is its ground truth. The straight line and the straight dashed line represent the BR-1 and BR-2, respectively. The action warning threshold ($\mathrm{y}_{\mathrm{fail}} = 1$) is illustrated by the horizontal dashed line and the current time is shown by the vertical dashed line. In the lower subfigures, the solid "bell-curve", dashed "bell-curve" and dotted "bell-curve" represent the failure-time distributions of the BR-1, BR-2 and ground truth, respectively.

It can be seen that BR-1 prediction for unit 1 (figure 3.5a) shows a better performance in contrast to BR-1 prediction for unit 11 (figure 3.5b). For unit 1, the choice of noise variance is appropriate as it is estimated from other units which also share the same noise variance across units. On the other hand, for unit 11, noise variance is estimated from other units which have different noise variance, therefore, this estimated



(a) (b)

Figure 3.6: The illustration of evolution of failure-time distributions of BR-1 and BR-2 for unit 1 (left) and unit 11 (right). At time index 0.13 where only few data are available, the failure-time distribution of BR-1 for unit 1 deviates slightly from the true failure-time distribution in contrast to failure-time distribution of BR-2 due to noise involvement, whereas a large deviation occurs for the failure-time distribution of BR-1 for unit 11 from the true failure-time distribution owing to high level of noise in degradation.

noise variance may not be accurate and cannot be changed over the time, because in BR-1, the noise variance is fixed.

Furthermore, figure 3.5a demonstrates that BR-1 and BR-2 predictions have "similar" performances, but BR-1 deviates slightly from the true value due to low level noise in degradation signal, whilst BR-2 can retain the prediction close to the true value, because the noise variance is also updated in each iteration, which take into account the variation in degradation. For the case when the noise variance vary and the high noise involved, such as shown in figure 3.5b, BR-1 prediction deviates from true value, whilst BR-2 prediction is able to cope with it due to its updated in noise variance.

Figure 3.6 illustrates the evolution of the predicted failure-time distributions for units 1 and 11 at time indexes 0.01, 0.13 and 0.77 for both approaches. BR-1 produces a failure-time distribution that is Bernstein distribution [Gebraeel et al., 2005], whilst BR-2 delivers an extension of the Bernstein distribution derived from the Student-t distribution (section 3.2.6). It can be seen that at time index 0.13, in figure 3.6a, where few data are available for unit 1, the failure-time distribution for BR-1 deviates slightly in comparison to the failure-time distribution for BR-2, because although the estimated

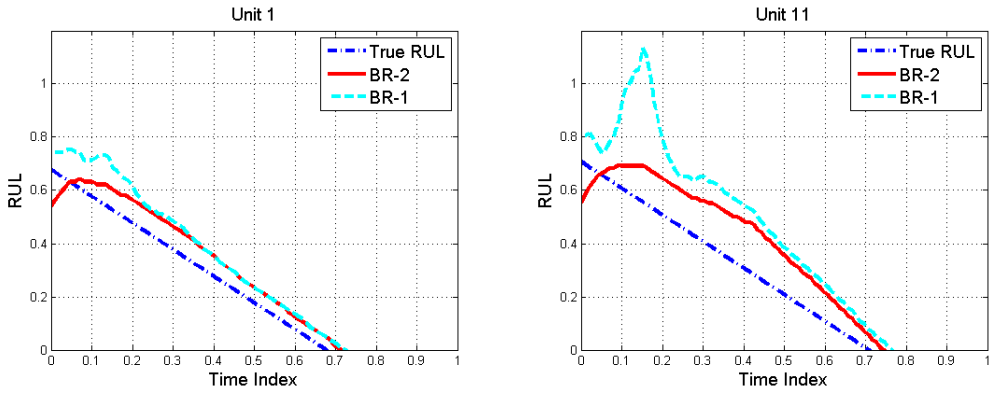
noise variance is appropriate, but the noise involved may still affect BR-1 performance. Figure 3.6b shows that BR-1 performance is poorer in comparison to BR-2, because the estimated noise variance in BR-1 may not be accurate and BR-1 cannot deal with high level of noise in degradation. At time index 0.77, after gathering enough data points, both failure-time distributions approach to the true failure-time distribution.

As these figures only provide a "snapshot" at a particular time, it is difficult to justify visually which approach has the better performance in general. It is therefore important to evaluate these approaches using a performance metric. There are two undesirable situations in prognostics: when the prediction is too early, called False Positive (FP) and when the prediction is too late, called False Negative (FN). RUL metric adopted from Goebel and Bonissone [2005]; Saxena et al. [2008a] is used to capture this.

In this simulation, RUL estimations are presented in figure 3.7, for units 1 and 11. The dot-dashed line is the actual RUL, the dashed and solid lines represent BR-1 and BR-2, respectively. As shown in figures 3.7a, there is minor deviation in both predictions to FN zone, whereas BR-2 is slightly better than BR-1. In figure 3.7b, the RUL estimations for BR-1 deviate largely from the true RUL in comparison to BR-2. For this scenario, BR-2 has better performance in terms of accuracy as well as consistency. The modelling noise variance as a random variable improves capability in Bayesian prognostics to capture large variation in the degradation signal.

In order to demonstrate a performance metric across all units, RUL residual is introduced. The RUL residual is the difference between the estimated RUL and the true RUL. A comparison of RUL residuals over the 10 synthetic datasets for each scenario can be illustrated by a time-indexed box plot, as shown in figures 3.8 (for units 1-10) and 3.9 (for units 11-20). On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers.

In figure 3.8, it is difficult to justify which approach is better. Once they have gathered enough data (approximately at time index 0.5), both box plots show similar performance. However, for the case when the noise variance vary between units, as shown in figure 3.9, the box plot for BR-1 shows more fluctuations indicating less consistency compared to BR-2. This figure also indicates that the median of the RUL residual across the time for BR-2 shows higher accuracy than BR-1. The result of this box plot strengthens the conclusion in previous figures that the inclusion of noise variance as a random



(a) (b)

Figure 3.7: The RUL comparisons between BR-1 and BR-2 using the synthetic data for units 1 (left) and 11 (right). For unit 1, there is a slight deviation in RUL for BR-1 compared to RUL for BR-2, because of the sensitivity of BR-1 to the noise. However, in general, both prognostics show "similar" performance, because the other data is generated from the same noise variance with low level of noise. On the other hand, the prognostics for unit 11 using BR-2 demonstrates performance increment than BR-1 when degradation signal has different variation between units with high level of noise.

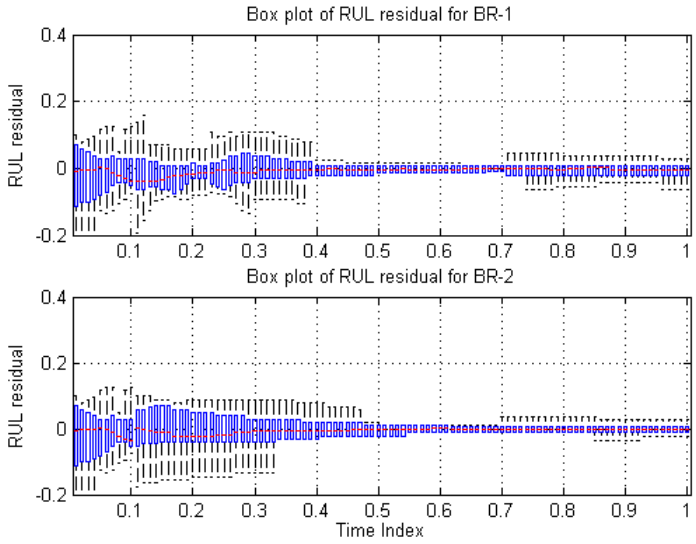


Figure 3.8: A comparison using box plots between BR-1 and BR-2 tested on 10 synthetic degradation data where the noise variance is kept the same across units. In general, the overall performance of BR-1 and BR-2 are "similar".

variable in the Bayesian model improves the quality of prediction in terms of accuracy and consistency.

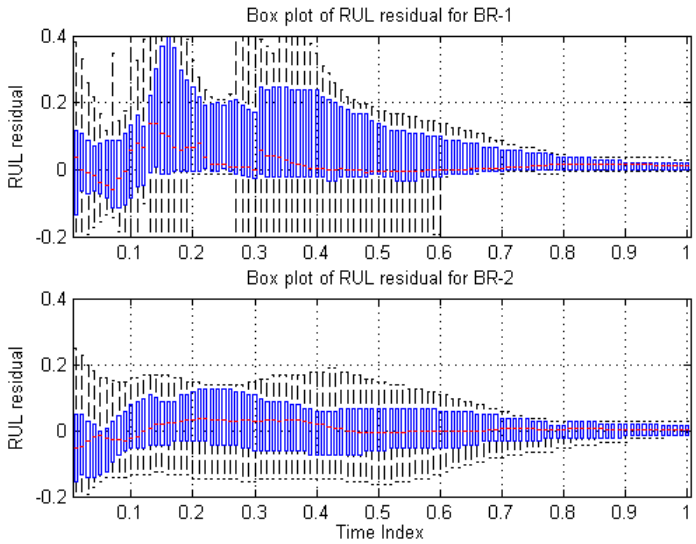
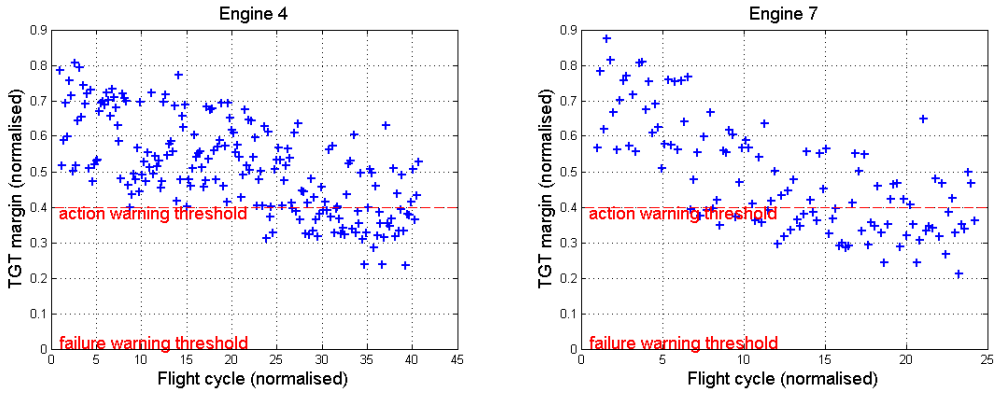


Figure 3.9: A comparison using box plots between BR-1 and BR-2 tested on 10 synthetic degradation data, where the noise vary between units with mixed noise (low and high levels of noise). The box plots represent the RUL residual across all units, where BR-2 demonstrates better capability in terms of accuracy and consistency in contrast to BR-1.

In conclusion, BR-1 works efficiently as a prognostic algorithm in two situations. First, when the noise variance across all units is the same (or at least similar). The second situation is when the degradation signal does not contain high level of noise. The noise variance is fixed and it cannot be updated, therefore the noise characteristics cannot be captured. This is the main reason why BR-1 was implemented successfully with bearing prognostics [Gebraeel, 2006; Gebraeel et al., 2005]. Despite the fact that health index in their application (RMS of vibration signal of bearing) may vary between units, it does not contain high level of noise and therefore the difference in noise variance can be neglected.

On the other hand, BR-2 is able to capture the noise variability across units and deals with high noise level in degradation, which is more likely to occur in reality. In gas turbine engine prognostics, TGT margin degradation data contains high level of noise, where BR-2 should deliver better RUL estimation than BR-1. To demonstrate



(a) (b)

Figure 3.10: Examples of normalised TGT margin degradation data for engines 4 (left) and 7 (right), where there is noise variation between these two degradations with high level of noise.

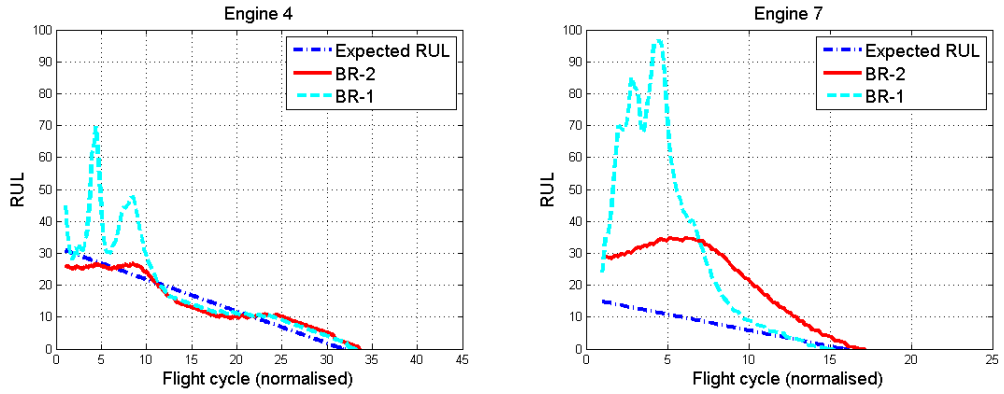
this, a real case study will be discussed in the following subsection.

3.3.2 Case study 2: Turbine Gas Temperature margin

This case study focuses on operational or in-service data; the take-off values of TGT margin for an aerospace gas turbine engine. This implementation aims to illustrate the effectiveness of the proposed algorithm using real degradation data. Ground truth RULs do not exist for this data. To estimate "expected RUL" at each flight cycle, linear regressions has been implemented on the complete sample, based on consultation with industrial peers. The predicted models are assumed to represent the true degradation in order to validate the prognostics. In this case, the action warning threshold is chosen as a set point to be 0.4.

Normalised TGT margin data has been supplied, measured from 10 aircraft engines. Engines 4 and 7 are selected arbitrarily for demonstration, shown by figure 3.10. On each figure, the symbol "+" represents TGT margin (normalised). It can be seen that the TGT margin degradation signals contain high level of noise and their noise vary between these two signals. The challenge in large uncertainty has also been described in section 1.2.3.3, where prognostic prediction might be inconsistent, especially when there is little data available.

For the real case study, the prognostic algorithms are applied to this data using the



(a) (b)

Figure 3.11: The RUL comparison between BR-1 and BR-2 using the real data for engine 4 (left) and engine 7 (right). RUL for BR-2 demonstrates better performance in terms of accuracy and consistency than RUL for BR-1.

same methodology as in case study 1. Figure 3.11 shows a comparison between BR-1 and BR-2 using RUL metric for engines 4 and 7. For both engines, RUL estimations using BR-1 fluctuate largely in comparison to BR-2, when there is little data available (approximately before flight cycle 10). BR-1 has poor performance because BR-1 cannot cope well with the noise variation between engines as well as the high level of noise in TGT margin degradation. In addition, figure 3.12 illustrates a comparison between BR-1 and BR-2 approaches tested on real degradation of gas turbine engine using box plots. For the first 16 flight cycles for all engines, the box plot for BR-1 fluctuates inconsistently and also has larger edges as well as longer whiskers in comparison to the box plot for BR-2. This emphasises the drawbacks of BR-1 in dealing with these prognostic issues. The figures 3.11 and 3.12 confirm that the synthetic data conclusions for BR-1 and BR-2 are borne out.

3.4 Summary

This chapter presents a Bayesian regression technique to predict degradation in a complex system that generalises existing methods [Gebraeel, 2006; Gebraeel et al., 2005]. Bayesian regression methods are promising for prognostic applications, delivering both an estimate of RUL with associated uncertainty bounds.

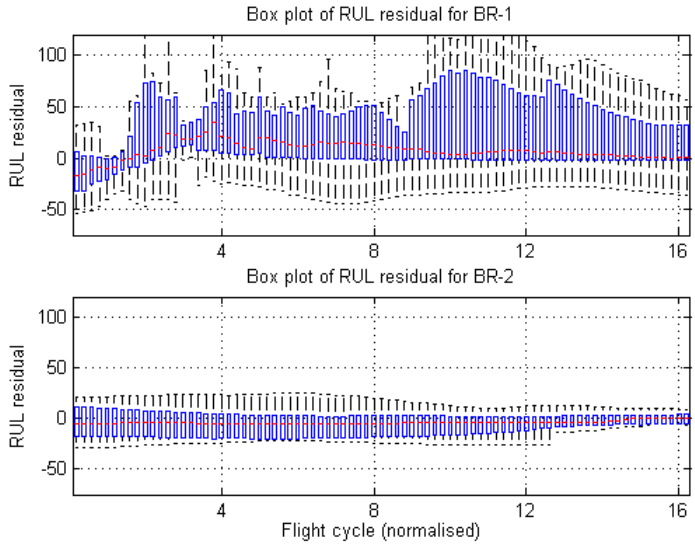


Figure 3.12: A comparison using box plots between BR-1 and BR-2 tested on the real data. The box plots represent the RUL residual across all units, where it demonstrates clearly that BR-2 is more consistent and accurate than BR-1.

BR-2 approach is introduced, which is fundamentally extended from BR-1 method. The main contribution in BR-2 lies in modelling of the observation noise as a random variable allowing for the variation in noise level. The strengths of BR-2 are investigated through the use of synthetic data. It is found that BR-2 is promising to be used when the degradation signals have different variation between units and they also have high level of noise. This situation is likely to occur in reality, such as TGT margin degradation problem. Therefore, in the second case study, BR-1 and BR-2 are applied to the real TGT margin data for validation. The noise included in the Bayesian model has shown the improvement of the results in terms of accuracy and consistency.

Another contribution of this work is a methodology for specifying prior distribution whose prior parameters can be determined directly from available degradation data. This results in full informative prior in Bayesian analysis. The importance of our methodology in specifying prior parameters will be analysed in a "more" complex Bayesian model[1] in the next chapter.

[1]The specification of prior parameters have a big impact in complex Bayesian model.

Chapter 4

Bayesian Hierarchical Model: a sampling method

4.1 Introduction

As discussed in section 1.2.3.4, there are abundant data available from multiple engines (the fleet), illustrated in figure 1.7. The state-of-the-art prognostic algorithms are able to update degradation models with data collected from individual assets with in-service measurements. However, the ability to accommodate the heterogeneous fleet into a single prognostic model is not well studied.

Both methods, BR-1 and BR-2, can be considered as Bayesian non-Hierarchical Models (BnHM). The main limitation of the prognostic approach using BnHM is that they do not utilise, optimally, data available from multiple (the fleet of) engines for estimating the RUL of one particular engine. Therefore, it is advantageous to maximise the value of gathered data by using a hierarchical modelling structure that is capable of accommodating multiple degradation signals gathered from same type of engines in a particular fleet. This type of model also allows to have more levels, e.g. for accommodating different types of engine or airline.

In the statistical literature [Gelman and Hill, 2006], this type of data is known as repeated measurement or panel data, which can be accommodated into a hierarchical model. Hierarchical models (also called multilevel models) are statistical models that have their own hierarchy, the model coefficients vary at more than one level [Gelman

and Hill, 2006]. This capability allows the models to obtain information from both prior distributions; an individual prior, related to a particular unit and a common prior, related to the whole fleet of units. This type of model has been extensively used in many fields to deal with panel data (econometrics), longitudinal data (biostatistics) and repeated measurement data (reliability).

In biostatistical applications, hierarchical models arise frequently in situations where several measurements are made on a number of individuals [Davidian and Giltinan, 1995]. In econometrics, the models are commonly used with panel data [Koop et al., 2007] that include understanding the diversity of preferences and sensitivities that exists in the market [Rossi et al., 2005] and in forecasting applications [Tobias, 2001].

In survival analysis and reliability engineering, hierarchical models can be used to create a natural model for failure count, failure time data and degradation data from multiple-unit system [Hamada et al., 2008; Ibrahim et al., 2001]. In addition, these models have been applied in political science [Gelman and Hill, 2006], spatial analysis [Banerjee et al., 2004] and many other statistical fields.

Hierarchical models have also been adopted recently in the field of machine learning. Examples of applications include computer vision [Fei-Fei and Perona, 2005], multimedia web page classification [Girolami and Rogers, 2005] and information retrieval [Blei et al., 2003]. However, the authors have not found any literature addressing about the use of hierarchical model in prognostic applications.

In this chapter, a Bayesian Hierarchical Model (BHM) is proposed, referred to as Bayesian Regression 3 (BR-3). A comparative study between the BR-2 and BR-3 approaches is presented. The algorithms are implemented on synthetic degradation data to evaluate their prognostic performance and subsequently, are applied on actual gas turbine in-service data.

4.2 Bayesian Regression 3

4.2.1 Model and block diagram

The BR-3 algorithm for prognostics is derived in this section. Lindley and Smith [1972] introduced a hierarchical model in the Bayesian framework. In complementary work, Gelfand et al. [1990] presented a solution for the posterior distribution for the hierarchi-

cal model using Markov Chain Monte Carlo (MCMC). Subsequently, this method has attracted many researchers to address growth curve analysis [Davidian and Giltinan, 1995; Menzefricke, 1999; Robinson and Crowder, 2000] which also covers the many areas mentioned above.

In the hierarchical model context, degradation signals obtained from each unit, $\mathbf{y}_j$, are modelled at the first model level, by:

$$\mathbf{y}_j = \mathbf{\Phi}_j \mathbf{w}_j + \boldsymbol{\varepsilon}_j \tag{4.1}$$

where j indexes the j^{th} unit, $\boldsymbol{\varepsilon}_j$ is a random error term that follows an iid Gaussian distribution, $\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, \sigma_j^2 I)$, and σ_j^2 is the noise variance, assumed to follow an IG distribution, $\sigma_j^2 \sim \text{IG}(a_j, b_j)$. $\mathbf{\Phi}_j$ is an $N_j \times D$ design matrix of basis function which maps an N_j dimensional input vector into a D dimensional feature space and $\mathbf{w}_j$ is a D dimensional vector of weights for j^{th} unit.

The second level of the model represents information common to the whole set of degradation signals. In order to express the commonality between units' parameters ($\mathbf{w}_j$), they are assumed to be drawn from a common Gaussian distribution. This can be written as:

$$p(\mathbf{w}_j) = \mathcal{N}(\bar{\mathbf{w}}, V) \tag{4.2}$$

where $\bar{\mathbf{w}}$ is the mean and V is the covariance of $\mathbf{w}_j$. The parameters $\bar{\mathbf{w}}$ and V^{-1} can be modelled as a Gaussian and a Wishart distribution, respectively; $\bar{\mathbf{w}} \sim \mathcal{N}(\eta, C)$ and $V^{-1} \sim \mathcal{W}([\rho R]^{-1}, \rho)$, where η is the mean of $\bar{\mathbf{w}}$, C is the covariance of $\bar{\mathbf{w}}$, R is a scale matrix and ρ denotes degree-of-freedom. The Wishart distribution is a conjugate prior for precision matrix of a multivariate Gaussian distribution [Bishop, 2006] and a multidimensional generalisation of the Gamma distribution [Murphy, 2007] (see also appendix A.4).

As for the BR-2 approach, this technique also combines two sources of information, the whole existing fleet data and real-time sensor information, to update, periodically, the RUL of individual units.

There are five main components in the proposed model, illustrated in figure 4.1. Components 1-3 are used to process the health signal. A degradation database contains historical degradation signals from many units. This information can be used to estimate prior parameters. The BR-3 approach is able to accommodate multiple degra-

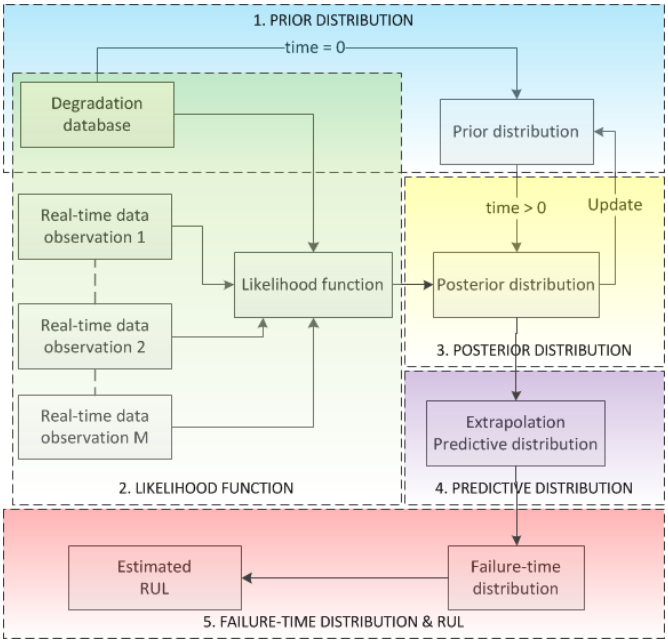


Figure 4.1: Block diagram of the BHM for prognostics. The main difference between BHM and BnHM, shown in figure 3.1, lies in accommodating multiple units in a single model.

dation signals monitored simultaneously from different units. These signals are used to compute the likelihood function. Having both the likelihood and the prior distribution, the posterior distribution can be calculated by Bayes's rule, equation (2.1), that is: posterior $\propto$ likelihood $\times$ prior.

The predictive distribution is then calculated and extrapolated over a desired time horizon (component 4). In the final step (component 5), the forecast is transformed into a failure time distribution to deliver an estimate of RUL.

4.2.2 Prior distribution

BR-3 has three prior distributions with six parameters.

$$p(\sigma_j^2|a_j, b_j) = \text{IG}(a_j, b_j) \tag{4.3}$$

$$p(\bar{\mathbf{w}}|\eta, C) = \mathcal{N}(\eta, C) \tag{4.4}$$

$$p(V^{-1}|\rho, R) = \mathcal{W}([\rho R]^{-1}, \rho) \tag{4.5}$$

The first prior is the *individual* prior, equation (4.3), whereas the last two priors, equations (4.4) and (4.5), are the *common* prior distributions.

BR-3 has more number of prior parameters than BR-2, leading to difficulty in obtaining highly informative prior. Therefore, various methods have been suggested to specify parameters for non-informative prior distributions. There is an extensive literature on the specification of diffuse priors for BHM, e.g. [Bernardo and Smith, 2009; Gelman et al., 2004; Kass and Wasserman, 1996; Spiegelhalter et al., 1996]. However, for this type of model, Gelman [2009] stated that non-informative distributions can sometimes have strong and undesirable implications.

In order to address this issue, in this subsection, a methodology to produce "moderately" informative prior distributions for BR-3 is proposed, where some of the prior parameters can be estimated from e.g. an in-service database.

First prior is IG distribution, $p(\sigma_j^2|a_j, b_j)$. In BHM, the choice of prior for the variances (e.g. IG distribution) is important and may be difficult [Daniels, 1999]. The most common non-informative prior choices for this distribution is Jeffreys prior [Jeffrey, 1961]. For IG family, Jeffrey's rule [Hamada et al., 2008; Jeffrey, 1961] defines a non-informative prior as:

$$p(\sigma^2) \propto 1/\sigma^2 \tag{4.6}$$

However, Gelman [2006] described serious problems with the IG family of non-informative prior distributions, where the resulting inferences may be sensitive.

A methodology to create highly informative IG prior is introduced. As described in section 3.2.2, the parameters a_j and b_j of equation (4.3) can be estimated by taking the squared residual values between predicted signal using OLS and the real measurement for each training degradation unit. Therefore, mean and variance of the squared residual values can be computed. Based on mean, $\hat{\mu}_j$, and variance, $\hat{\sigma}_j^2$, formulation for IG distribution for each unit j, the parameters, a_j and b_j, can be calculated by:

$$a_j = \left(\frac{\hat{\mu}_j}{\hat{\sigma}_j}\right)^2 + 2 \tag{4.7}$$

$$b_j = \hat{\mu}_j \left(\left(\frac{\hat{\mu}_j}{\hat{\sigma}_j}\right)^2 + 1\right) \tag{4.8}$$

where for a new unit, these parameters can be initiated by computing the mean of train-

Table 4.1: The methodology of estimating R_{sf} parameter for Wishart distribution in BR-3.

R_{sf}	V^{-1}	Effects
"small"	"large"	Common
"large"	"small"	Individual

ing prior parameters, a_j and b_j.

The next prior is $\bar{\mathbf{w}} \sim \mathcal{N}(\eta, C)$. The parameter, η, can be estimated by calculating the mean of all the $\mathbf{w}_j$ obtained by OLS. To express the prior uncertainty about the value of this common mean, the parameter C can be selected to be diffuse.

Finally, the prior distribution of V^{-1} has two parameters, ρ and R. The parameter, ρ, is restricted to $\rho > D - 1$ to ensure that the Gamma function in the normalisation factor is well-defined (appendix A.4). The parameter, R, can be determined by calculating covariance of the OLS estimation of $\mathbf{w}_j$.

The scaling of R (R_{sf}) can be determined based on table 4.1. By increasing the R_{sf}, the prediction tends to follow the individual, such as a standard Bayesian regression (e.g. BR-2). However, the common prior will play a dominant role in prediction by specifying R_{sf} to be small [Tobias, 2001].

4.2.3 Likelihood function

Multiple observed degradation signals are used to compute the likelihood function. The likelihood for our model is conditional on the joint probability of observing the data ($\mathbf{X}_j$) and the model parameters ($\mathbf{w}_j, \sigma_j^2$), written:

$$p(\mathbf{y}_j|\mathbf{X}_j, \mathbf{w}_j, \sigma_j^2) = \mathcal{N}(\mathbf{\Phi}_j \mathbf{w}_j, \sigma_j^2 I) \tag{4.9}$$

where, j is the unit number.

4.2.4 Posterior distribution

Having the likelihood function and the prior distributions, a joint posterior distribution for all the parameters of this model can be computed using Bayes' theorem by:

$$p(\mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2 | \mathbf{y}_j) \propto \left[\prod_{j=1}^{M} p(\mathbf{y}_j | \mathbf{X}_j, \mathbf{w}_j, \sigma_j^2)\, p(\mathbf{w}_j | \bar{\mathbf{w}}, V^{-1})\, p(\sigma_j^2 | a_j, b_j) \right] p(\bar{\mathbf{w}} | \eta, C)\, p(V^{-1} | \rho, R) \tag{4.10}$$

where M is the total number of units.

The joint posterior distribution has a complex form where the expectation is not analytically tractable. In this case, a sampling method is required to compute the posterior distribution.

The conjugacy of the specifications in equations (4.2)-(4.5) at each stage in the hierarchy delivers simple complete conditional posterior densities for all parameters of interest, where this makes implementation of the Gibbs sampler convenient.

In order to obtain a posterior distribution for each parameter, it is necessary to derive their posterior conditional distributions [Gelfand et al., 1990; Koop et al., 2007; Menzefricke, 1999] (see also appendix C.1 for the detailed derivation). Each complete posterior conditional is proportional to the aforementioned joint posterior. Thus, all of the terms in the product of equation (4.10) not involved in each posterior conditional are absorbed into the normalising constant of this conditional. The detail of the complete posterior conditionals will be discussed next.

4.2.4.1 Posterior conditional for $\mathbf{w}_j$

All of the terms in the product of equation (4.10) that do not involve $\mathbf{w}_j$ are absorbed into the normalising constant of this conditional. Hence, posterior conditional for $\mathbf{w}_j$ can be calculated by:

$$\begin{aligned} p(\mathbf{w}_j | \mathbf{X}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2, \mathbf{y}_j) &\propto p(\mathbf{y}_j | \mathbf{X}_j, \mathbf{w}_j, \sigma_j^2) p(\mathbf{w}_j | \bar{\mathbf{w}}, V) \\ p(\mathbf{w}_j | \mathbf{X}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2, \mathbf{y}_j) &= \mathcal{N}(d_{\mathbf{w}_j}, D_{\mathbf{w}_j}) \end{aligned} \tag{4.11}$$

where

$$D_{\mathbf{w}_j} = \left(\mathbf{\Phi}_j^T \mathbf{\Phi}_j / \sigma_j^2 + V^{-1}\right)^{-1} \tag{4.12}$$

$$d_{\mathbf{w}_j} = D_{\mathbf{w}_j} \left(\mathbf{\Phi}_j^T \mathbf{y}_j / \sigma_j^2 + V^{-1} \bar{\mathbf{w}}\right) \tag{4.13}$$

4.2.4.2 Posterior conditional for $\bar{\mathbf{w}}$

All of the terms in the product of equation (4.10) that do not involve $\bar{\mathbf{w}}$ are absorbed into the normalising constant of this conditional. Hence, posterior conditional for $\bar{\mathbf{w}}$ can be calculated by:

$$\begin{aligned} p(\bar{\mathbf{w}}|\mathbf{X}_j, \mathbf{w}_j, V^{-1}, \sigma_j^2, \mathbf{y}_j) &\propto \left[\prod_{j=1}^{M} p(\mathbf{w}_j|\bar{\mathbf{w}}, V^{-1})\right] p(\bar{\mathbf{w}}|\eta, C) \\ p(\bar{\mathbf{w}}|\mathbf{X}_j, \mathbf{w}_j, V^{-1}, \sigma_j^2, \mathbf{y}_j) &= \mathcal{N}(d_{\bar{\mathbf{w}}}, D_{\bar{\mathbf{w}}}) \end{aligned} \tag{4.14}$$

where

$$D_{\bar{\mathbf{w}}} = \left(MV^{-1} + C^{-1}\right)^{-1} \tag{4.15}$$

$$d_{\bar{\mathbf{w}}} = D_{\bar{\mathbf{w}}} \left(MV^{-1} \left(\frac{1}{M} \sum_{j=1}^{M} \mathbf{w}_j\right) + C^{-1}\eta\right) \tag{4.16}$$

4.2.4.3 Posterior conditional for σ_j^2

All of the terms in the product of equation (4.10) that do not involve σ_j^2 are absorbed into the normalising constant of this conditional. Hence, posterior conditional for σ_j^2 can be calculated by:

$$\begin{aligned} p(\sigma_j^2|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \mathbf{y}_j) &\propto p(\mathbf{y}_j|\mathbf{X}_j, \mathbf{w}_j, \sigma_j^2) p(\sigma_j^2|a_j, b_j) \\ p(\sigma_j^2|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \mathbf{y}_j) &= \text{IG}(\hat{a}_j, \hat{b}_j) \end{aligned} \tag{4.17}$$

where

$$\hat{a}_j = \frac{N_j \times M}{2} + a_j \tag{4.18}$$

$$\hat{b}_j = \left[\frac{1}{2}(\mathbf{y}_j - \mathbf{\Phi}_j \mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j \mathbf{w}_j) + b_j^{-1}\right]^{-1} \tag{4.19}$$

where N_j is the sample size for each unit.

4.2.4.4 Posterior conditional for V^{-1}

All of the terms in the product of equation (4.10) that do not involve V^{-1} are absorbed into the normalising constant of this conditional. Hence, posterior conditional for V^{-1} can be calculated by:

$$p(V^{-1}|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, \sigma_j^2, \mathbf{y}_j) \propto \left[\prod_{j=1}^{M} p(\mathbf{w}_j|\bar{\mathbf{w}}, V^{-1})\right] p(V^{-1}|\rho, R)$$

$$p(V^{-1}|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, \sigma_j^2, \mathbf{y}_j) = \mathcal{W}([\hat{\rho}\,\hat{R}]^{-1}, \hat{\rho}) \tag{4.20}$$

where

$$\hat{\rho} = M + \rho \tag{4.21}$$

$$\hat{R} = \left[\sum_{j=1}^{M}(\mathbf{w}_j - \bar{\mathbf{w}})(\mathbf{w}_j - \bar{\mathbf{w}})^T + \rho R\right]^{-1} \tag{4.22}$$

The samples from the posterior conditional distributions in equations (4.11), (4.14), (4.17) and (4.20) can be generated to implement a Gibbs sampler [Gelfand et al., 1990; Gelfand and Smith, 1990; Koop et al., 2007]. This yields a set of G_s draws, from Gibbs sampler, given by:

$$\left\{\mathbf{w}_j^{(g_s)}, \bar{\mathbf{w}}^{(g_s)}, (\sigma_j^2)^{(g_s)}, (V^{-1})^{(g_s)}\right\}, g_s = 1, \cdots, G_s \tag{4.23}$$

In general, the idea behind Gibbs sampler is to generate a sequence of draws which, after a suitable "pre-convergence"[1], have converged in distribution to the joint posterior density. The post-convergence draws (draws kept after the burn-in) can then be used

[1]Pre-convergence is the number of initial iteration that should be discarded to allow for "burn-in".

to estimate posterior statistical properties, such as mean, standard deviation or other quantities of interest [Tobias, 2001].

After removing the first B iteration of these samples, i.e. burn-in sample, the remaining $G_s - B$ draws can be averaged to create estimates of posterior features of interest. Therefore, the mean, the standard deviation or other statistical properties of the samples from the posterior conditionals of $\mathbf{w}_j, \bar{\mathbf{w}}, \sigma_j^2$ and V^{-1} can be calculated. In the following subsection, the algorithm 1 describes pseudocode of Gibbs sampler algorithm.

4.2.5 Predictive distribution

The predictive density of the future of a degradation signal $\mathbf{y}_j^*$ can be obtained by marginalising the conditional predictive density over the posterior $p(\mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2|\mathbf{y}_j)$:

$$p(\mathbf{y}_j^*|\mathbf{X}_j^*, \mathbf{y}_j) = \int p(\mathbf{\Phi}_j^*\mathbf{w}_j, \sigma_j^2|\mathbf{y}_j)p(\mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2|\mathbf{y}_j)\,\mathrm{d}\mathbf{w}_j\,\mathrm{d}\bar{\mathbf{w}}\,\mathrm{d}V^{-1}\,\mathrm{d}\sigma^2 \quad (4.24)$$

In this case, the required integration does not have a closed-form analytical solution, while the dimensionality of the space and the complexity of the integrand may prohibit numerical integration based on one-dimensional integral methods[1].

Draws from the posterior predictive density for $\mathbf{y}_j^*$ can be obtained using post-burn-in draws from the Gibbs sampler. This performs direct Monte Carlo integration using equation (4.24) where draws from $p(\mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \sigma^2|\mathbf{y}_j)$ are taken from the Gibbs sequence after it has been determined to converge. A draw from the marginal posterior predictive density of $\mathbf{y}_j^*$ can be obtained by taking draws from the following normal density:

$$(\mathbf{y}_j^*)^{(g_s)} \sim \mathcal{N}\left(\mathbf{\Phi}_j^*\mathbf{w}_j^{(g_s)}, (\sigma_j^2 I)^{(g_s)}\right) \quad (4.25)$$

The algorithm 1 describes pseudocode for Gibbs sampler solution for posterior conditional distributions and Monte Carlo integration for predictive distribution. Based on these steps, the mean and standard deviation of the sample of predictive distribution of $\mathbf{y}_j^{*(g_s)}$ can be estimated to produce trajectory prediction.

[1]The examples include rectangle rule, trapezoidal rule, etc.

Algorithm 1 Gibbs sampler for posterior conditional $\{\mathbf{w}_j, \bar{\mathbf{w}}, \sigma_j^2, V^{-1}\}$ and Monte Carlo integration for predictive distribution $\mathbf{y}_j^*$

Specify starting values $\{\bar{\mathbf{w}}^0, (\boldsymbol{\sigma}^{\mathbf{2}})^0, (V^{-1})^0\}$.
for $g_s = 0, \cdots, (G_s - 1)$ **do**
  for $j = 1, \cdots, M$ **do**
    Draw $\mathbf{w}_j^{(g_s+1)} \sim p(\mathbf{w}_j|\mathbf{X}_j, \bar{\mathbf{w}}^{(g_s)}, (V^{-1})^{(g_s)}, (\boldsymbol{\sigma}^{\mathbf{2}})^{(g_s)}, \mathbf{y}_j)$, eq. (4.11).
    $\mathbf{W}(:, j)^{(g_s+1)} = \mathbf{w}_j^{(g_s+1)}$
    Draw $(\sigma_j^2)^{(g_s+1)} \sim p(\sigma_j^2|\mathbf{X}_j, \mathbf{W}^{(g_s+1)}, \bar{\mathbf{w}}^{(g_s)}, (V^{-1})^{(g_s)}, \mathbf{y}_j)$, eq. (4.17).
    $\boldsymbol{\sigma}^{\mathbf{2}}(:, j)^{(g_s+1)} = (\sigma_j^2)^{(g_s+1)}$
  end for
  Draw $\bar{\mathbf{w}}^{(g_s+1)} \sim p(\bar{\mathbf{w}}|\mathbf{X}_j, \mathbf{W}^{(g_s+1)}, (V^{-1})^{(g_s)}, (\boldsymbol{\sigma}^{\mathbf{2}})^{(g_s+1)}, \mathbf{y}_j)$, eq. (4.14).
  Draw $(V^{-1})^{(g_s+1)} \sim p(V^{-1}|\mathbf{X}_j, \mathbf{W}^{(g_s+1)}, \bar{\mathbf{w}}^{(g_s+1)}, (\boldsymbol{\sigma}^{\mathbf{2}})^{(g_s+1)}, \mathbf{y}_j)$, eq. (4.20).
  Draw $(\mathbf{y}_j^*)^{(g_s+1)} \sim p(\mathbf{y}_j^*|\mathbf{X}_j^*, \mathbf{y}_j) = \mathcal{N}(\boldsymbol{\Phi}_j^* \mathbf{w}_j^{(g_s+1)}, (\sigma_j^2 I)^{(g_s+1)})$, eq. (4.25).
end for
Calculate the mean of $\mathbf{y}_j^*$: $\mathbb{E}(\mathbf{y}_j^*) \approx \frac{1}{(G_s - B)} \sum_{g_s=B}^{G_s-1} [(\mathbf{y}_j^*)^{(g_s+1)}]$.
Calculate the variance of $\mathbf{y}_j^*$: $\text{var}(\mathbf{y}_j^*) \approx \frac{1}{(G_s - B)} \sum_{g_s=B}^{G_s-1} [(\mathbf{y}_j^*)^{(g_s+1)} - \mathbb{E}(\mathbf{y}_j^*)]^2$.

4.2.6 Failure-time distribution and RUL

In this case, the failure-time distribution $f(t)$ can be calculated using a Bernstein distribution [Ahmad and Sheikh, 1984; Gebraeel et al., 2005], because the predictive distribution is drawn from a Gaussian distribution, equation (4.25). The probability of the event that the estimated failure time (T) exceeds time (t) is equivalent to the probability that the future degradation signal ($\mathbf{y}_j^*$) is less than a chosen failure threshold ($\mathbf{y}_{\text{fail}}$). This can be written as:

$$P(T > t) = P(\mathrm{y}_j^* \leq \mathrm{y}_{\text{fail}}) \tag{4.26}$$

Therefore, the failure-time distribution can be formulated as:

$$\begin{aligned} F(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P(\mathrm{y}_j^* \leq \mathbf{y}_{\text{fail}}) \\ &= 1 - \Phi\left(\frac{\mathrm{y}_{\text{fail}}^* - \boldsymbol{\Phi}_j^* \mathbf{w}_j}{\sigma_j}\right) \end{aligned} \tag{4.27}$$

$$f(t) = \frac{dF(t)}{dt} \tag{4.28}$$

where $\Phi(.)$ is the cumulative distribution function of a Gaussian distribution. The parameters of $\mathbf{w}_j$ and σ_j are obtained from the draws of predictive distribution, equation (4.25).

As described in section 3.2.6, RUL can be estimated by subtracting the current time from the mode of the failure-time distribution, $f(t)$, known as the estimated failure-time, denoted by T. The current time is denoted by τ. So, the RUL is then given by: $\text{RUL} = T - \tau$.

4.2.7 Algorithm summary

In summary, figure 4.2 shows the graphical models for BR-2 (left) and BR-3 (right). The main difference between these models can be seen clearly in the parameters $\bar{\mathbf{w}}$ and V. In BR-3, these parameters are modelled as random variables following Gaussian and Wishart distributions, respectively, while in BR-2, these parameters are treated as fixed parameters. In addition, BR-3 also accommodates multiple units M with different data samples, shown by the dashed box. Symbol N_j is the data sample size and M is the total number of units.

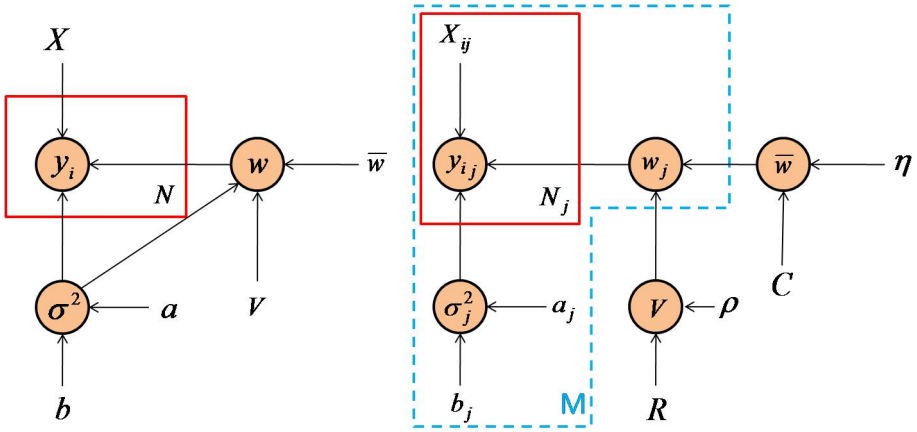


Figure 4.2: The differences between BR-2 and BR-3 in the form of graphical model. BR-3 adopts an additional level of hierarchy by treating the parameters $\bar{\mathbf{w}}$ and V as random variables, while in BR-2, these parameters are assumed fixed. The dashed and solid boxes illustrate that BR-3 can accommodate multiple units with different data samples.

Figure 4.3 illustrates a block diagram of a complete methodology for BR-3 ap-

proach. Prior distributions are estimated from degradation database. All degradation units together with degradation database are fed into BR-3 algorithm, where the details of BR-3 block diagram are shown in figure 4.1. RUL for all units can be estimated simultaneously, even if they have different sample size of data. This process will continue until a maintenance action decision is made (or when the observation complete). For each completed degradation cycle, the data will be added to degradation database to update new prior distributions. For a new unit, it can be fed directly into BR-3 to estimate its RUL.

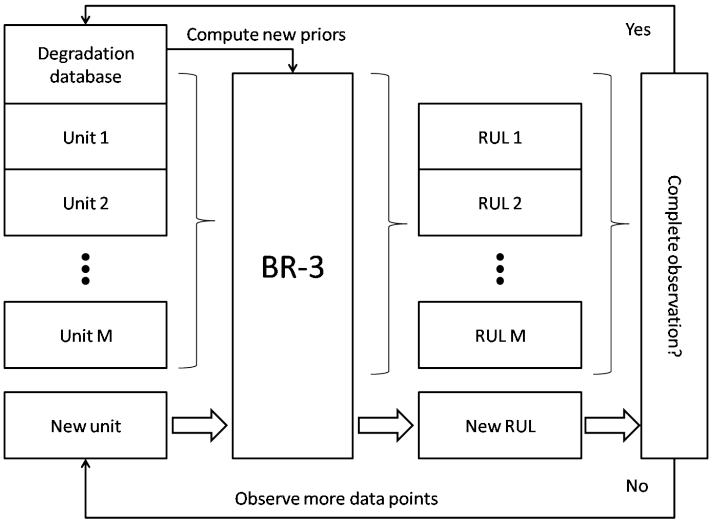


Figure 4.3: An illustration of block diagram of complete BR-3 approach. Having estimated prior parameters for all units from degradation database, multiple units with different data samples are fed into a a single BR-3 algorithm, where RUL can be estimated simultaneously. For a new unit, it can be fed directly into BR-3 to estimate its RUL.

4.3 Case Studies

In chapter 3, the advantages of using BR-2 have been demonstrated for dealing with standard TGT margin degradation[1]. In this section, the benefits of using BR-3 are demonstrated, where it can accommodate heterogeneous fleet (section 1.2.3.4) to deal

[1]Here, we define standard degradation as degradation data with linear pattern which occur in most of obtained TGT margin degradation.

with more complicated challenges associated with complex systems, including large uncertainty in degradation (section 1.2.3.3) as well as multiple degradation pattern (section 1.2.3.5).

The comparison of two prognostic algorithms, BR-2 and BR-3, will be presented. As in section 3.3, the algorithms are tested on two case studies: synthetic degradation data and actual service data exhibiting gas turbine engine degradation. The proposed method for testing and comparing the prognostic algorithms begins by testing them on synthetic degradation data to gain confidence in their applicability. Subsequently, the prognostic approaches are applied to the real TGT margin data to confirm their capability.

4.3.1 Case Study 1: Synthetic degradation data

The synthetic data are generated to model the real data characteristics. In this case, a standard degradation signal is generated as described in section 3.3.1, using equation (3.18) and two extreme scenarios are produced to characterise the real-world situation, such as degradations with non-linearity and very high noise level.

First extreme scenario is, where the rates of degradation may be non-linear. In this case, we generate synthetic degradation data, $\mathbf{d}_3$, based on a piecewise linear model, given by:

$$\mathbf{d}_3 = \begin{cases} \boldsymbol{\Phi}_a \mathbf{w}_a + \boldsymbol{\varepsilon}_a & \text{if } x_a \leq x_c, \\ [\mathbf{1}_a \, \mathbf{x}_a]\mathbf{w}_a + \boldsymbol{\varepsilon}_a & \\ & \\ \boldsymbol{\Phi}_b \mathbf{w}_b + \boldsymbol{\varepsilon}_b & \text{if } x_b > x_c, \\ [\mathbf{1}_b \, \mathbf{x}_b]\mathbf{w}_b + \boldsymbol{\varepsilon}_b & \end{cases} \tag{4.29}$$

where $\boldsymbol{\varepsilon}_a \sim \mathcal{N}(0, \sigma_a^2 I)$ and $\boldsymbol{\varepsilon}_b \sim \mathcal{N}(0, \sigma_b^2 I)$ and they vary from one unit to another unit. Model coefficients are represented by $\mathbf{w}_a$ and $\mathbf{w}_b$, where they are drawn from Gaussian distribution, $\mathbf{w}_a \sim \mathcal{N}(\mu_{\mathbf{w}_a}, \Sigma_{\mathbf{w}_a})$ and $\mathbf{w}_b \sim \mathcal{N}(\mu_{\mathbf{w}_b}, \Sigma_{\mathbf{w}_b})$. These can be decomposed as: $\mathbf{w}_a = [w_{a1} \; w_{a2}]^T$, $\mathbf{w}_b = [w_{b1} \; w_{b2}]^T$. Parameters $\mathbf{1}$ and $\mathbf{x}$ are N dimensional column vector of 1s and input variable (time index), respectively, where $\mathbf{1} = [\mathbf{1}_a \; \mathbf{1}_b]^T$, $\mathbf{x} = [\mathbf{x}_a \; \mathbf{x}_b]^T$, $x_a \, \epsilon \, \mathbf{x}_a$ and $x_b \, \epsilon \, \mathbf{x}_b$. x_c is a time index where the change in degradation's slope

occurs.

In order to make this scenario more realistic[1], it is necessary to set $w_{b2} < w_{a2}$, whereas the continuity of piecewise-defined functions can be achieved by applying limit theorem[2] to find w_{b1} based on other model parameters $\{w_{a_1}, w_{a_2}, w_{b2}\}$ and x_c.

Another extreme scenario is when degradation contains very high level of noise[3]. In this case, synthetic degradation data are generated from a linear model with very high noise level, given by:

$$\mathbf{d}_4 = \mathbf{\Phi}_4 \mathbf{w} + \boldsymbol{\varepsilon}_4 \tag{4.30}$$

where $\mathbf{w} \sim \mathcal{N}(\mu_{\mathbf{w}}, \Sigma_{\mathbf{w}})$ and $\boldsymbol{\varepsilon}_4 \sim \mathcal{N}(0, \sigma_4^2 I)$, the noise variance also varies from unit to unit. For this case, $\mathbf{d}_4$, first-order polynomial basis function is used, so that $\mathbf{\Phi} = [\mathbf{1}\ \mathbf{x}]$. To produce very high noise level in degradation signal, it is necessary to set: $\sigma_4^2 > \{\sigma_2^2, \sigma_a^2, \sigma_b^2\}$.

Fourteen degradation units are generated based on similarity with the real degradation behaviour, as shown in figure 4.4, where four units characterise extreme scenarios: each two are generated from equations (4.29) and (4.30). The remaining units are generated from equations (3.18) to represent standard degradation signal. Time index is used to represent the number of flight cycles (normalised), and health index embodies the engine's health to be forecast. In this case, the action warning threshold is chosen to be 22.

The signals for unit 2 and 11 are selected for prognostics because they represent the extreme behaviour of generated scenarios. Unit 2 represents a degradation signal with a sudden rate change behaviour, shown by the symbol "+" and its solid line illustrates the ground truth. It can be seen that the signal decreases gradually until time index around 0.25, after that, it starts to degrade rapidly. Unit 11 and its ground truth, illustrated by the symbol "○" and solid line, respectively. It has high noise which occurs at the beginning of the degradation signal, highlighted by an oval region. The other degradation signals are shown by dashed lines.

[1]This scenario is when the rate of degradation slope changes after time index x_c, where in reality this might occur due to step change in covariate or fault, see section 1.2.3.5.

[2]See theories of functions, limits and continuity.

[3]In reality, this may occur owing to variation in operating condition or environmental factor, see section 1.2.3.3.

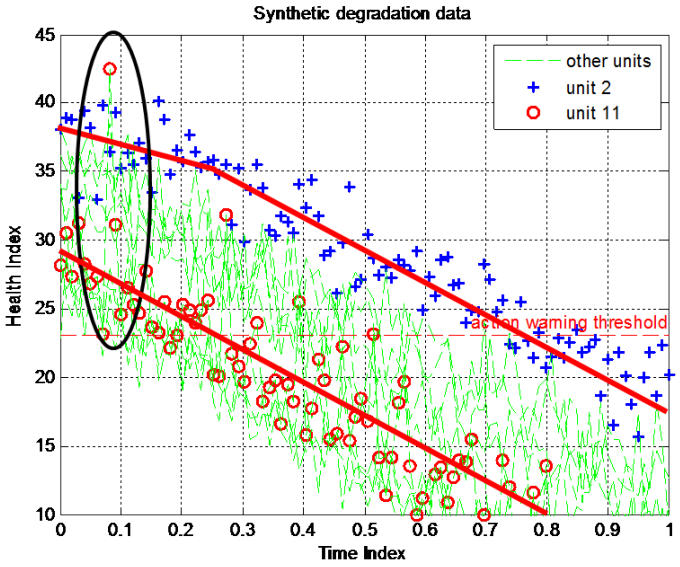


Figure 4.4: The synthetic data for 14 units. Owing to their extreme nature, the signals for unit 2 (symbol "+") and 11 (symbol "○") are selected for prognostics.

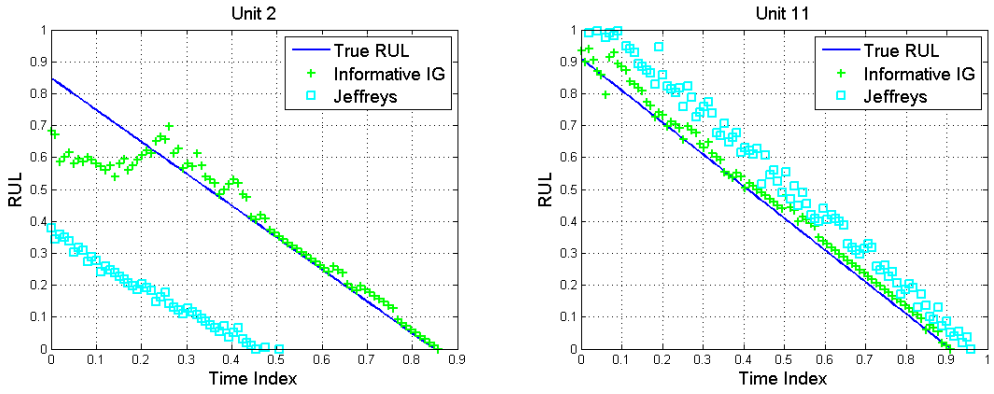
As described in section 4.2.2, the prior specification is crucial in BHM. Therefore, the proposed methodology for prior specification will be analysed first in this following subsection, before the benefits of BR-3 in prognostics are investigated.

4.3.1.1 Informative IG prior vs Jeffreys prior

This subsection presents the advantage using "moderately" informative prior, contains our proposed highly informative IG prior, in comparison to a well-known non-informative prior for BHM, a Jeffreys prior.

The prior parameters are obtained based on the same procedure which is described in section 4.2.2. Some parameters are selected to be diffuse, including C, which is covariance of prior $\bar{\mathbf{w}}$, and ρ, which is a degree-of-freedom parameter in Wishart prior, V^{-1}. To express the prior uncertainty about the value of this common mean, the parameter C can be selected to be five times larger than its mean. This produces a substantially flat prior over each element [Tobias, 2001]. Furthermore, the parameter, ρ, is set to five to ensure propriety of the Wishart prior [Koop et al., 2007; Tobias, 2001].

The parameter, R, can be determined by calculating covariance of the OLS estimation of $\mathbf{w}_j$, and here the R_{sf} is set to be one. For the proposed IG prior distribution,



(a) Unit 2 (b) Unit 11

Figure 4.5: RUL comparison between BR-3 with informative IG prior and BR-3 with Jeffreys prior. The results reveal that the use of Jeffreys prior might produce unexpected inference.

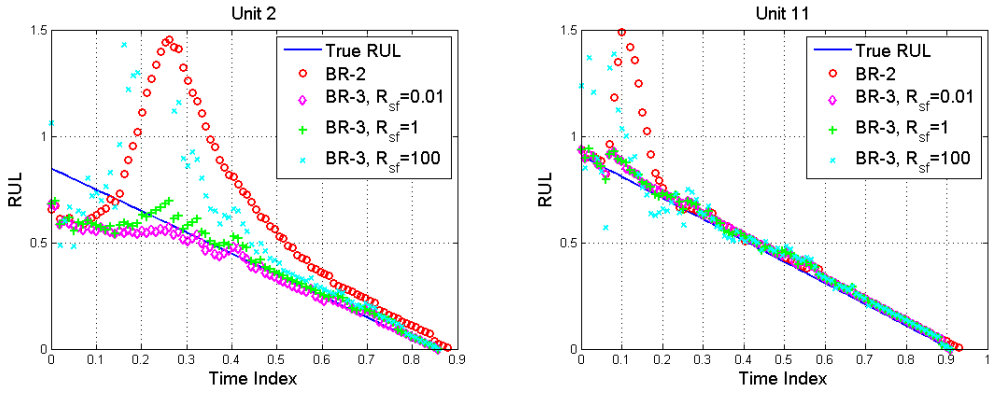
parameters a_j and b_j can be calculated based on equations (4.7) and (4.8), whilst for Jeffreys prior, the equation (4.5) can be used by setting $p(\sigma^2) = \text{IG}(0.001, 0.001)$ [Lesaffre and Lawson, 2012].

In this case, two extreme scenarios are used to demonstrate the comparison of two determining prior parameters. The iteration for Gibbs sampling is set to be $G_s = 10,000$, with burn-in time period is set to be $B = 1,000$. As in previous chapter, here we still use first-order polynomial basis function.

Figure 4.5 shows two RUL estimations for units 2 and 11. Solid line is true RUL and symbols "+" and "□" represent RUL for BR-3 with informative IG prior and Jeffreys prior, respectively. The results demonstrate clearly that the use of Jeffreys prior leads to poor performance in RUL estimation. The utilisation of Jeffreys prior may require numerous data points to converge to the true RUL as shown in figure 4.5b. Therefore, the proposed methodology to determine crucial prior parameters, e.g. IG, is recommended to be used in BR-3 for prognostics. From now onward, this setting of prior parameters will be used for BR-3 prognostics.

4.3.1.2 The advantages of BR-3 in accommodating heterogeneous fleet

This subsection discusses the effect of selecting scaling factor of prior parameter R (R_{sf}) and demonstrates the advantages of BR-3 in dealing with extreme scenarios that



(a) Unit 2 (b) Unit 11

Figure 4.6: RUL comparison between BR-2 and BR-3 using the synthetic data for unit 2 and 11. The choice of R_{sf} is suggested to be 1, therefore BR-3 has a moderate common prior (symbol "+") and it is able to deal well with extreme prognostic problems.

may occur in complex system degradation.

The same setting of prior parameters and Gibbs sampling are used as described in previous subsection, and units 2 and 11 are also selected for demonstration. Three simulations are generated using different values of R_{sf} (scaling factor of R), including $R_{sf} = 100$, $R_{sf} = 1$ and $R_{sf} = 0.01$. As described in section 4.2.2, the choice of this scaling factor will influence the behaviour of BR-3, whether it behaves as BHM or BnHM. This study aims to provide the appropriate choice of R_{sf} for prognostics and to demonstrate the strength of BR-3 for the extreme scenarios.

Figure 4.6 demonstrates RUL metrics for units 2 and 11. The solid line is the true RUL, symbols "○" embodies RUL for BR-2, whereas symbols "⋄", "+" and "×" represent RUL estimations for BR-3 with $R_{sf} = 0.01$, $R_{sf} = 1$ and $R_{sf} = 100$, respectively.

Owing to the effect of a sudden rate change in degradation in unit 2, BR-2 deviates strongly from the true value (figure 4.6a), whilst in unit 11, the BR-2 also diverges from the true value because of the high noise level early in the degradation (figure 4.6b). This occurs because BR-2 relies solely on an individual prior. As described in table 4.1, RUL estimation using BR-3 with $R_{sf} = 100$ also tends to follow the individual, e.g. BR-2.

In contrast, by specifying R_{sf} to be small, e.g. $R_{sf} = 1$ and $R_{sf} = 0.01$, they are able to retain the RUL estimations close to the true RUL because common priors (level 2) play an important role in the prediction. The prediction can share mutual health

information across the units, and combine it with the individual prior information (level 1). Both types of prior in BR-3 approach are able to enhance the RUL estimation. It is suggested to select R_{sf} to be "moderate", e.g. $R_{sf} = 1$, because in figure 4.6a, BR-3 with $R_{sf} = 0.01$ shows slower convergence than BR-3 with $R_{sf} = 1$ due to a very strong common prior for $R_{sf} = 0.01$. This would be an issue when the initial prediction is far from the true value. Therefore, from now onward, it is necessary to set $R_{sf} = 1$ for BR-3 prognostics.

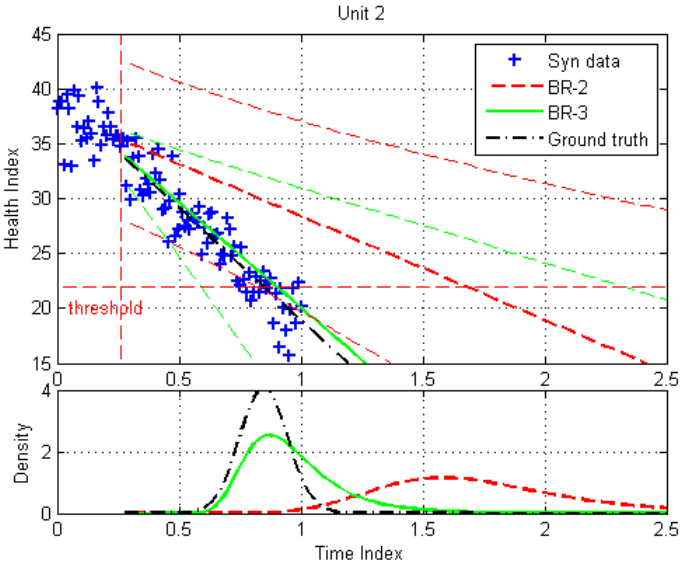


Figure 4.7: Trajectory (degradation) predictions of synthetic data for unit 2 at time index 0.25. The BR-2 prediction deviates from the true value, whilst the BR-3 retains the prediction close to the truth, when there is a sudden change behaviour in the degradation signal.

Figure 4.7 demonstrates the trajectory (degradation) prediction of synthetic data for unit 2 at time index 0.25 onward, i.e. inference is performed only on data for unit 2, up to time index 0.25. In the top subfigure, the straight dashed line and the straight solid line represent the BR-2 and BR-3 predictions, respectively. The symbol "+" is the degradation value corrupted by noise whereas the straight dot-dashed line is its ground truth. In this figure, the action warning threshold ($\mathbf{y}_{\text{fail}} = 22$) is illustrated by the horizontal dashed line and current time is shown by the vertical dashed line. In the lower subfigure, the dashed, solid and dot-dashed "bell-curves" represent the failure-time distribution of BR-2, BR-3 and ground truth, respectively. This figure illustrates that when

the "gradual decrease" in degradation signal is only available, the BR-2 prediction deviates strongly from the truth, being dominated by the initial shallow slope of unit 2, whereas, the BR-3 retains the prediction close to the truth because of information sharing across units permitted by the additional hierarchical level.

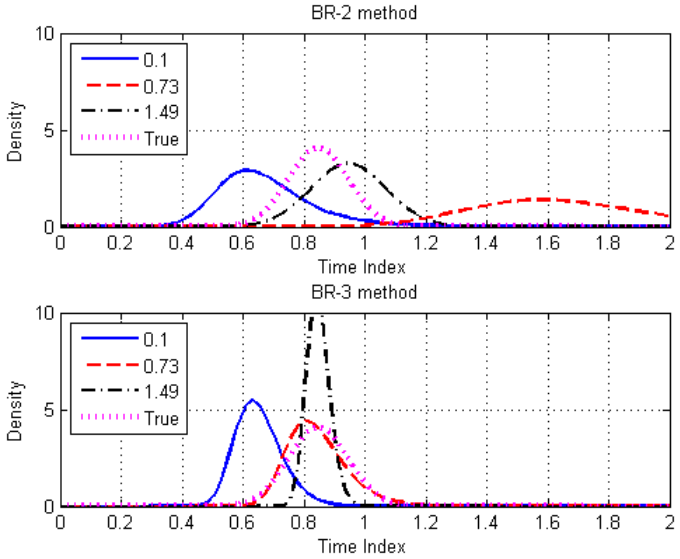
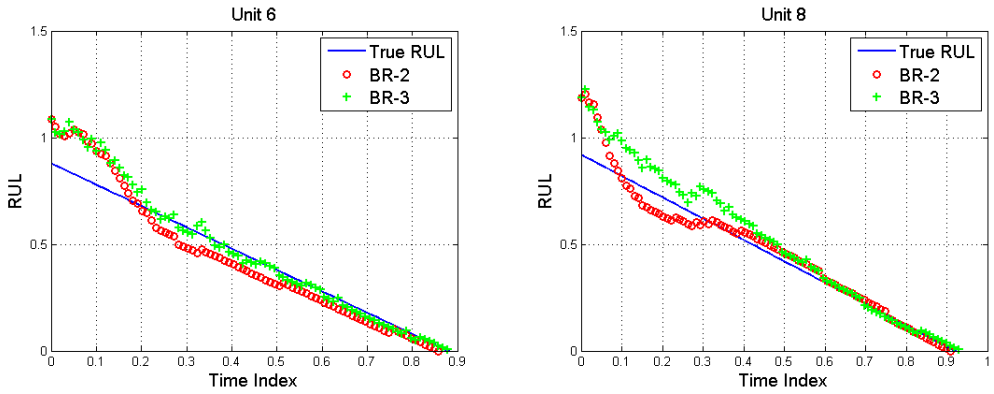


Figure 4.8: The evolution of failure-time distributions of BR-2 and BR-3. At time index 0.73, failure-time distribution of BR-2 deviates strongly from the truth in comparison to failure-time distribution of BR-3. Common priors in BR-3 share information from other units to influence BR-3 prediction about how degradation may occur in the future.

Figure 4.8 illustrates the evolution of predicted failure-time distributions for unit 2 at time indexes 0.1, 0.73 and 1.49 for both approaches. The dotted "bell-curve" is the true failure-time distribution. It can be seen that failure-time distribution of BR-3 converges consistently to true mode of failure-time distribution as time progresses, whilst the failure-time distribution for BR-2 deviates significantly at time index 0.73. However, when enough data are gathered, both failure-time distributions approach close to the truth. As described in section 4.2.6, BR-3 failure-time distribution is a Bernstein distribution, whereas the failure-time distributions of BR-2 is an extension of Bernstein distribution, described in section 3.2.6.

For the case of standard degradation units, there is no substantial difference between BR-2 and BR-3 as shown in figure 4.9. This figure shows RUL for units 6 and 8, where they are referred to standard degradation units.



(a) Unit 6 (b) Unit 8

Figure 4.9: RUL comparison between BR-2 and BR-3 using the synthetic data for standard degradation (units 6 and 8). For this case, there is no substantial difference between BR-2 and BR-3.

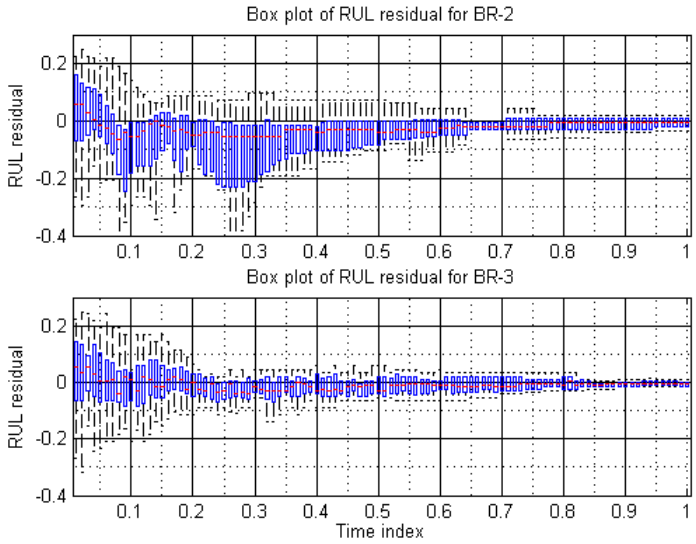
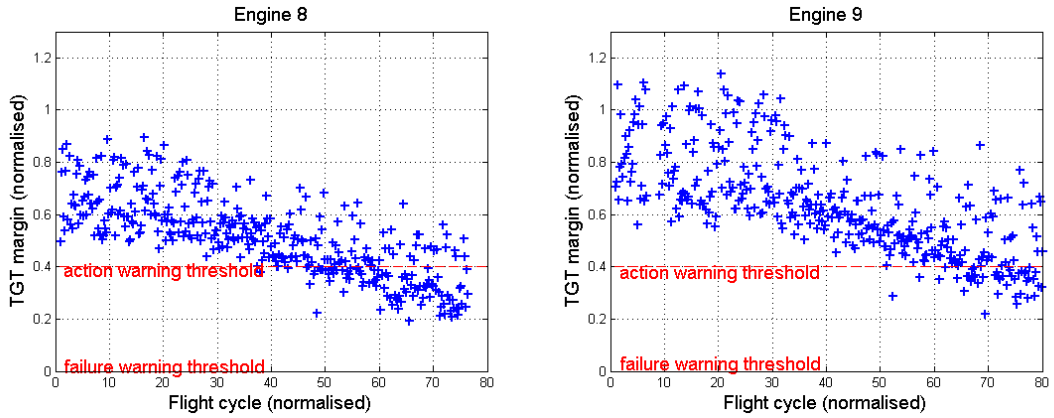


Figure 4.10: Box plot representing RUL performance for all synthetic data. Due to the involvement of some extreme prognostic scenarios, BR-2 shows more overall variability, indicates lack of consistency and accuracy for overall prognostic performance.



(a) Engine 8 (b) Engine 9

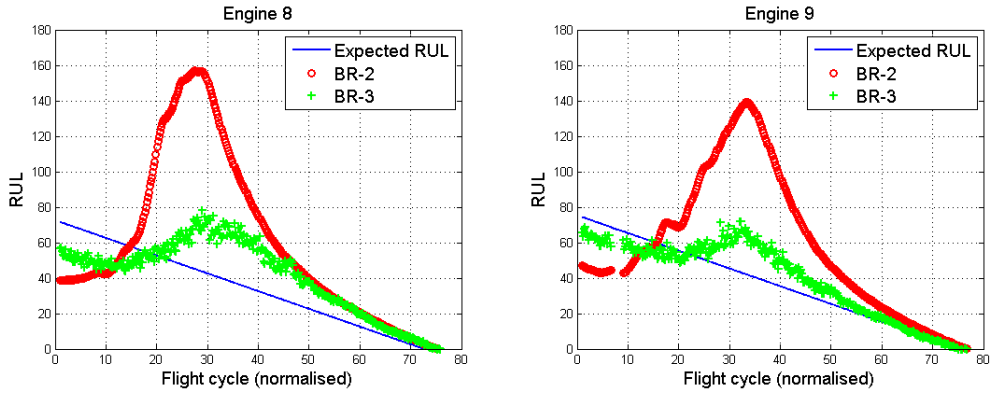
Figure 4.11: Examples of normalised TGT margin degradation. Engine 8 is more likely to has a change in the degradation slope whereas engine 9 has very high noise.

To summarise, performance across all units, the RUL residual is calculated for each individual at each time in turn. The RUL residual is the difference between estimated RUL and true RUL. A comparison of RUL residuals for all data is illustrated by a time-indexed box plot in figure 4.10. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers. The box plot for BR-2 shows more overall variability, which means lack of consistency and accuracy, especially in early stages before time index of 0.5, in comparison to BR-3. This occurs because in this case study, there are some extreme prognostic scenarios involved. The box plot also emphasises that BR-3 is more suitable as a prognostic algorithm, where in reality, unexpected extreme scenarios may take place.

4.3.2 Case Study 2: Turbine Gas Temperature margin

This case study focuses on in-service data, the take-off values of TGT margin for an aerospace gas turbine engine. As described in section 3.3.2, this section aims to illustrate the effectiveness of the proposed algorithm using real degradation data. In this case, the action warning threshold is still chosen as a set point to be 0.4.

In this case study, there are 14 TGT margin data, generated from 14 aircraft engines. Figure 4.11 shows two examples of take-off TGT margin degradation data (normalised).



(a) Engine 8 (b) Engine 9

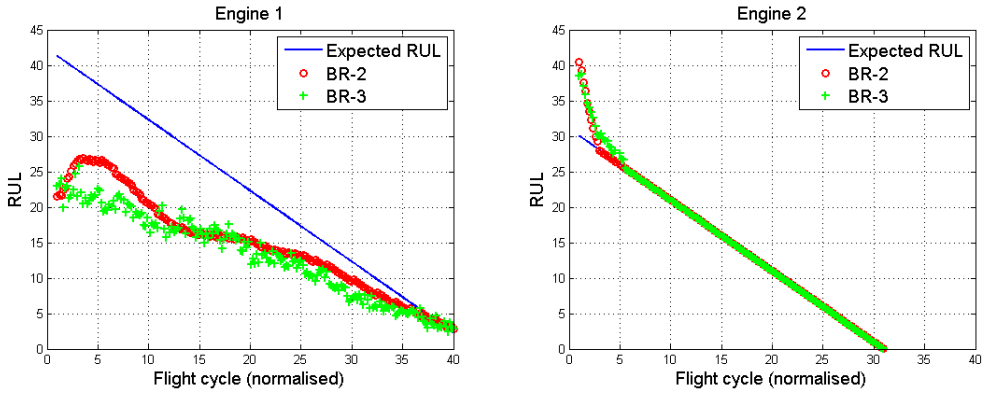
Figure 4.12: RUL comparison between BR-2 and BR-3 using real TGT margin data for engines 8 and 9. BR-3 can cope well with the extreme scenarios in prognostics (slope change and very high noise in degradation).

The x-axis represents the time index, which is the normalised number of flight (cycle) whereas the y-axis embodies the health index which is the normalised TGT margin.

Engines 8 and 9 are selected for prognostics because engine 8 is more likely to have a sudden change in degradation behaviour, illustrated in figure 4.11a and engine 9, because there is very high noise level in the signal, shown in figure 4.11b.

The prognostic algorithms are applied to these data using the same methodology as in case study 1. Figure 4.12 shows the comparisons between BR-2 and BR-3 using RUL metric for the engines 8 and 9. The solid line is the "expected RUL", the symbols "$\bigcirc$" and "$+$" represent RUL for BR-2 and BR-3, respectively. Figure 4.12a reveals that there is a significant deviation in RUL for BR-2 due to the sudden change in degradation behaviour, whereas RUL for BR-3 can still maintain the prediction close to the "expected RUL" value. A similar result is also shown in figure 4.12b: the BR-2 prediction deviates largely in comparison to BR-3 from the "expected RUL" due to very high noise level in TGT margin data, whilst the BR-3 is still able to deal with the high noise level in TGT margin data. After receiving enough data, both predictions converge to the "expected RUL" value. The conclusions drawn from the synthetic data for BR-2 and BR-3 carry over to this case study: a Bayesian approach involving two types of prior (BR-3) in both levels, is able to enhance RUL estimation for these extreme scenarios.

Furthermore, figure 4.13 shows RUL for engines 1 and 2 where they are considered



(a) Engine 1 (b) Engine 2

Figure 4.13: RUL comparison between BR-2 and BR-3 using real TGT margin data for engines 1 and 2, considered as standard degradation data. There is no substantial difference for the RUL between BR-2 and BR-3.

as "standard" degradation signals. It can be seen that for this case, RULs for BR-2 and BR-3 do not have substantial difference.

Figure 4.14 shows a box plot, representing RUL residual of BR-2 and BR-3 when applied to all available real degradation data. Because in reality, the total number of flight cycles varies from engine-to-engine, we show only the first 28 here. It can be observed that box plot for BR-2 has longer whiskers. Also at the end of flight cycle (approximately from flight cycle 18 onward), the box plot for BR-2 has larger edges than BR-3. This indicates that BR-2 is less accurate and consistent. Figure 4.14 proves that the conclusions drawn from the previous synthetic data are accurate.

4.4 Summary

This chapter presents a comparative study between two advanced Bayesian prognostic approaches; Bayesian non-Hierarchical Model (BnHM) and Bayesian Hierarchical Model (BHM), named BR-2 and BR-3, respectively. These approaches are implemented on synthetic degradation data to evaluate their relative performances. Subsequently, we find that these characteristics are preserved when applied to actual service data, exhibiting gas turbine engine health degradation.

Civil aerospace is able to generate a considerable number of health signals from a

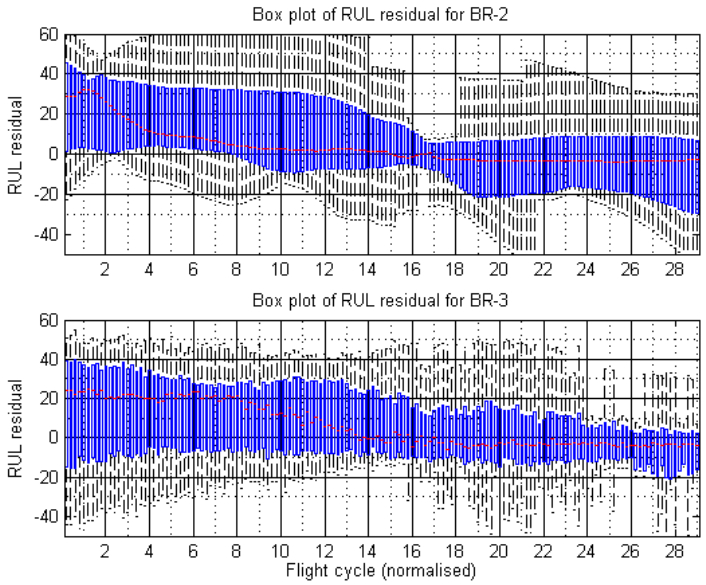


Figure 4.14: Boxplot representing RUL performance for all real data. Box plot for BR-3 shows smaller edges and shorter whiskers than box plot for BR-2, indicate that BR-3 is more consistent and accurate in prognostic analysis involving some extreme scenarios.

fleet of assets. BR-3 prognostics is appropriate to maximise the use of the assets' data by accommodating them into a hierarchical model. In this way, the health information can be shared between the engines in order to estimate robustly the RUL of the engines.

First contribution here is a methodology for specifying parameters of prior distributions using "moderately" informative prior, where the some parameters of the noise variance can be determined directly from the available degradation data. This methodology in specifying IG parameters, is compared with a well-know prior choice for IG distribution, named Jeffreys prior. The results show that for this application, our methodology improves the BR-3 performance and reduces the number of uninformative priors in the BR-3 method.

Another contribution is in demonstrating the strengths of BR-3 in prognostic problems. In general, there is no substantial difference between BR-3 and BR-2, applied to standard degradation data. However, BR-3 outperforms BR-2 for extreme situations in degradation signal. To demonstrate this, two engine degradation signals, one with a sudden degradation rate change, and the other with very high noise levels, are selected. The

results reveal that BR-3 has better performance than BR-2 for these extreme scenarios. BR-3 is able to share the mutual health information due to the common priors (level 2) and this is then combined with individual prior information (level 1). Therefore, BR-3 is promising to be applied in complex system prognostics, where there may be some extreme scenarios, occurring in degradation signals.

In term of practicability, BR-3 may have some drawbacks owing to the use of sampling method (Gibbs sampler). In the next chapter, these drawbacks will be discussed and a proposed solution to substitute sampling method will be introduced.

Chapter 5

Bayesian Hierarchical Model: a variational approach

5.1 Introduction

As described in chapter 4, BR-3 does not have a closed-form solution, which requires the use of the Gibbs sampler for computing conditional posterior distributions. Although, Gibbs sampler has emerged as an extremely popular tool for the analysis of complex statistical models, it suffers from well-known and potentially serious drawbacks from a practical point of view. It can be extremely computationally demanding, even for relatively small-scale statistical problems [Barber, 2012]. In addition, it can be difficult to determine the length of the required "pre-convergence" [Hjort et al., 2010] and to decide when it is safe to terminate the sampling and conclude its "convergences[1]" [Winn and Bishop, 2005]. That is, at what point it is reasonable to believe that the samples are truly representative of the underlying stationary distribution of the Markov chain [Cowles and Carlin, 1996]. An expert in Monte Carlo methods, may therefore need to be involved for determining the total number of iteration required as well as "pre-convergence" period[2].

BR-3 with Gibbs sampler also suffers from another drawback where the sampling should be restarted in each prognostic iteration, consequently inefficiency would take

[1]Convergences is sufficient run length of a simulation.
[2]This process is called diagnostic convergence.

place. A sequential type of Markov Chain Monte Carlo (MCMC) may address this problem, but it would require high level of expertise to run this.

In practice, such issues will result in less autonomy in the prognostic system where a statistician is still required to determine the critical configurations based on expert knowledge. This reduces the quality of prognostic function itself, e.g. eliminating human intervention in equipment monitoring. Final complication arises when BR-3 with sampling method uses abundant data, e.g. collected from tens of thousands of engines, computational capacity might also be an issue. A big investment is required by implementing a powerful super computer with parallel computing capabilities. This realisation may be extremely expensive and ultimately it might not generate prognostic benefits in term of return on investment.

These drawbacks become the main motivations to explore the possibility of using approximate inference for BR-3. Highly efficient deterministic approximation schemes have been recently developed [Bishop, 2006]. These methods offer a complementary alternative to sampling methods and have allowed Bayesian techniques to be used in large-scale applications [Blei et al., 2003]. This chapter introduces variational Bayes, as an alternative to Gibbs sampler methods, for taking a fully Bayesian approach to statistical inference over complex distributions that are difficult to directly evaluate or sample from, such as those that arise in BR-3.

5.2 Bayesian Regression 3 using variational inference

A central task in the application of Bayesian models is the evaluation of the posterior distribution. However, for complex Bayesian models, such as BR-3, evaluating of the posterior distribution is not analytically tractable. In this case, it is necessary to resort to approximation schemes: the stochastic, such as Gibbs sampling as described in chapter 4, or the deterministic, such as Laplace approximation, expectation propagation and variational Bayes.

Laplace approximation aims to find a Gaussian approximation to a probability density defined over a set of continuous variables. However, the main limitation of this framework is that it is based on a local Gaussian approximation to a mode (i.e., a maximum) of the distribution and so can fail to capture important global properties [Bishop, 2006].

Expectation propagation is another alternative form of deterministic approximate inference [Minka, 2001a,b]. The principle is similar with variational Bayes, but it is in the reverse form. Winn and Bishop [2005] stated that expectation propagation is limited to certain classes of model for which the required expectations can be evaluated and it is also not guaranteed to converge in general.

Since BR-3 uses a product of well-known probability distributions (exponential family distributions with conditional conjugacy), it is known that variational Bayes can provide a good approximation to the true posterior distribution [Beal, 2003; Bishop, 2006]. This section will focus on variational Bayes for BR-3, based on mean-field approximation[1] [Beal, 2003; Bishop, 2006].

5.2.1 Variational inference

The key objective of variational method is to approximate the integral with a simpler form that is tractable, forming a lower or upper bound. The integration then translates into the simpler problem (i.e. in term of implementability) of bound optimisation: making the bound as tight as possible to the true value [Beal, 2003].

Now let consider in more detail how the concept of variational optimisation can be applied to the inference problem in BR-3. BR-3 has latent variables[2], θ, and an observed variable[3], Y. Therefore, the joint distribution is $p(Y, \theta)$[4] and the goal is to calculate an approximation for the posterior distribution $p(\theta|Y)$ as well as the marginal distribution $p(Y)$. The log marginal probability can be decomposed as:

$$\ln p(Y) = \mathcal{L}(q) + \mathrm{KL}(q||p) \tag{5.1}$$

where

$$\mathcal{L}(q) = \int q(\theta) \ln \left\{ \frac{p(Y, \theta)}{q(\theta)} \right\} d\theta \tag{5.2}$$

$$\mathrm{KL}(q||p) = -\int q(\theta) \ln \left\{ \frac{p(\theta|Y)}{q(\theta)} \right\} d\theta \tag{5.3}$$

[1] The mean-field approximation is the case in which each approximated distribution is fully factorised.

[2] A latent variable is a hidden variable, such as the model parameters.

[3] In our case study later, an example of observed variable is TGT margin data.

[4] We omit the dependence on model choice for notational simplicity, but note that this is applicable throughout the following chapter.

where $q(\theta)$ is approximate distributions, $\mathrm{KL}(q||p)$ is the Kullback-Leibler divergence between $q(\theta)$ and the posterior distribution $p(\theta|Y)$. The Kullback-Leibler divergence satisfies $\mathrm{KL}(q||p) \geq 0$, with equality if, and only if, $q(\theta) = p(\theta|Y)$ [Kullback and Leibler, 1951]. Therefore, it follows from equation (5.1) that $\mathcal{L}(q) \leq \ln p(Y)$, in other words that $\mathcal{L}(q)$ is a lower bound on $\ln p(Y)$. The detailed verification of equation (5.1) can be found in appendix D.1.

The lower bound $\mathcal{L}(q)$ can be maximised by optimisation with respect to the distribution $q(\theta)$, which is equivalent to minimising the KL divergence. If any possible choice for $q(\theta)$ is allowed, then the maximum of the lower bound occurs when the KL divergence vanishes, which occurs when $q(\theta)$ equals the posterior distribution $p(\theta|Y)$.

In order to achieve a good approximation, it is important to restrict the family of distributions $q(\theta)$, so that they comprise only tractable distributions as well as allowing the family to be sufficiently rich and flexible [Bishop, 2006; Winn and Bishop, 2005]. One method to restrict the family of distributions $q(\theta)$ is to assume that they are independent [Jaakkola, 2001; Jordan et al., 1999]. Then, the elements of θ can be partitioned into disjoint groups that can be denoted by θ_g where $g = 1, \cdots, G$. Hence the q distribution can be factorised by:

$$q(\theta) = \prod_{g=1}^{G} q_g(\theta_g) \tag{5.4}$$

Next, equation (5.4) is substituted into equation (5.2), and then dissect out the dependence on one of the factors $q_h(\theta_h)$. Therefore, a general solution for optimal factor, $q_h^*(\theta_h)$, is obtained, given by:

$$\ln q_h^*(\theta_h) = \mathbb{E}_{g \neq h}[\ln p(Y, \theta)] + \text{const} \tag{5.5}$$

where const is a constant and $\mathbb{E}_{g \neq h}$ is the expectation with respect to variables present in the remaining factors $g \neq h$. In other words, for each factor, the log of the joint distribution is taken over all variables and then averages with respect to those variables not in that factor. The detailed verification of equation (5.5) is described in appendix D.1.

5.2.2 Model and block diagram

BR-1 and BR-2 are examples of BnHM, described in section 2.4.2 and chapter 3. As described in equations (2.5) and (3.1), the model can be expressed as:

$$\mathbf{y} = \mathbf{\Phi w} + \boldsymbol{\varepsilon} \tag{5.6}$$

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$. In BR-1, noise variance parameter, σ^2, is treated as a fixed variable, whilst in BR-2, it is assumed as a random variable, following an inverse Gamma distribution, $\sigma^2 \sim \text{IG}(a, b)$.

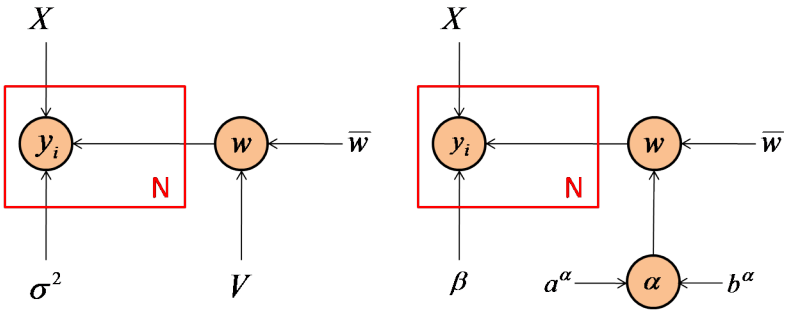


Figure 5.1: A comparison of graphical models for BR-1 based on analytical (left) and variational inference (right).

Bishop [2006] and Drugowitsch [2008] described how BR-1 and BR-2, which have deterministic solutions, can be solved approximately using variational inference. Figures 5.1 and 5.2 compare graphical models for analytical and variational inference for BR-1 and BR-2, respectively. Precision is preferred to be used in variational Bayes, for simplicity in computation, so that $\beta^{-1} = \sigma^2$. In order to ensure that the variational solution will be analytically tractable [Bishop, 2006], another additional hierarchical prior is introduced over precision of the model's weight, for BR-1 and BR-2. The conjugate prior for the precision of a Gaussian is a Gamma distribution, so that $\alpha \sim \text{Gam}(a^\alpha, b^\alpha)$.

However, there are no real benefits from applying variational inference for BR-1 and BR-2, because they are already analytically tractable. On the other hand, since BR-3 does not have a closed-form solution, applying variational inference will deliver benefits to deal with the drawbacks of the sampling method.

A similar principle of variational inference for BR-1 and BR-2, is adopted for BR-3. For simplicity in computation of variational Bayes, precision is used in preference to

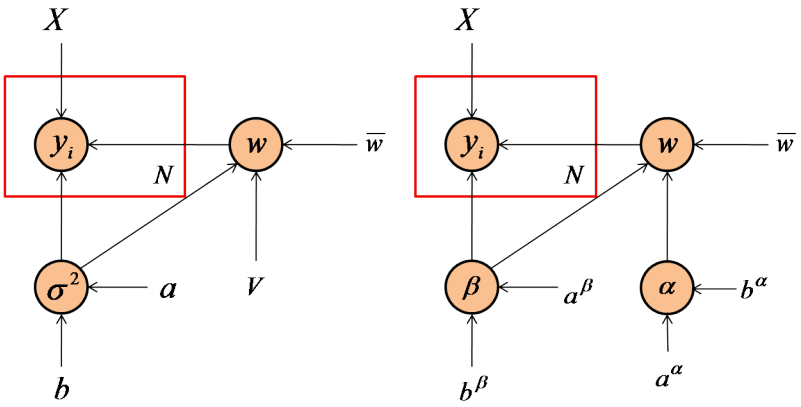


Figure 5.2: A comparison of graphical models for BR-2 based on analytical (left) and variational inference (right).

covariance, so that the equivalent symbols are $\sigma_j^2 = \beta_j^{-1}$ and $V = \Lambda^{-1}$. The same BR-3 model as described in chapter 4, equation (4.1), is used. Thus, the degradation signals obtained from each engine, $\mathbf{y}_j$, are modelled at the first hierarchical level, given by:

$$\mathbf{y}_j = \mathbf{\Phi}_j \mathbf{w}_j + \boldsymbol{\varepsilon}_j \tag{5.7}$$

where j indexes the j^{th} unit, $\boldsymbol{\varepsilon}_j$ is a random error term that follows an iid Gaussian distribution, $\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, \beta_j^{-1} I)$, and β_j is the noise precision, assumed to follow a Gamma distribution, $\beta_j \sim \text{Gam}(a_j^\beta, b_j^\beta)$. $\mathbf{\Phi}_j$ is a design matrix of basis functions which maps an N_j dimensional input vector into a D dimensional feature space and $\mathbf{w}_j$ is a D dimensional vector of weights.

The second level of the model represents information common to the whole set of degradation signals. In order to express the commonality between units' parameters ($\mathbf{w}_j$), they are assumed to be drawn from a common Gaussian distribution. This can be written as:

$$p(\mathbf{w}_j) = \mathcal{N}(\bar{\mathbf{w}}, (\beta_j \Lambda)^{-1}) \tag{5.8}$$

where $\bar{\mathbf{w}}$ is the mean and $(\beta_j \Lambda)^{-1}$ is the covariance of $\mathbf{w}_j$. $\bar{\mathbf{w}}$ is D dimensional vector and Λ is a $D \times D$ positive-definite matrix. The parameters $\bar{\mathbf{w}}$ and Λ can be assumed to follow a Gaussian and a Wishart distribution, respectively; $\bar{\mathbf{w}} \sim \mathcal{N}(\eta, (\alpha\Lambda)^{-1})$ and $\Lambda \sim \mathcal{W}([\rho R]^{-1}, \rho)$. Where η is the mean of $\bar{\mathbf{w}}$, $(\alpha\Lambda)^{-1}$ is the covariance of $\bar{\mathbf{w}}$, whilst R is a scale matrix and ρ denotes the degree-of-freedom of the Wishart distribution of Λ.

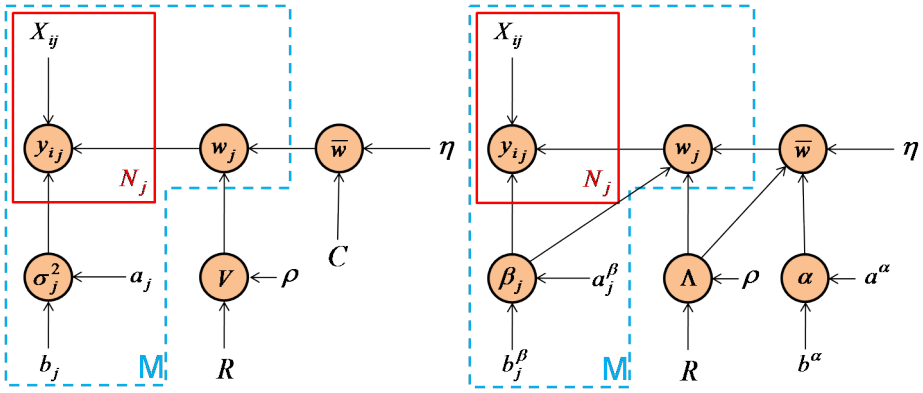


Figure 5.3: A comparison of graphical models for BR-3 using Gibbs sampler (left) and Variational Bayes (right).

Figure 5.3 shows a comparison of the two graphical models; BR-3 with Gibbs sampler and variational solution on the left and right hand side, respectively. It can be seen that the main differences are that a further hierarchical prior is introduced and some edges are connected to ensure that the variational posterior distribution can be computed analytically, assuming they are chosen to be conjugate-exponential [Fox and Roberts, 2012]. This figure also shows that both models accommodate multiple units M with different data samples, shown by the dashed box. Symbol N_j is the data sample size and M is the total number of units.

5.2.3 Prior distributions

Unlike BR-3 using Gibbs sampler, BR-3 using variational Bayes (BR-3-VB) has four prior distributions to be specified, owing to the introduction of the hierarchical prior for precision, as shown in figure 5.3. These are:

$$p(\beta_j) = \text{Gam}(\beta_j | a_j^\beta, b_j^\beta) \tag{5.9}$$

$$p(\bar{\mathbf{w}}|\Lambda) = \mathcal{N}(\bar{\mathbf{w}}|\eta, (\alpha\Lambda)^{-1}) \tag{5.10}$$

$$p(\Lambda) = \mathcal{W}(\Lambda|\rho, R) \tag{5.11}$$

$$p(\alpha) = \text{Gam}(\alpha | a^\alpha, b^\alpha) \tag{5.12}$$

The first distribution is an *individual* prior, equation (5.9), and the remaining, equations (5.10)-(5.12), are *common* priors. These parameters can be estimated from e.g. an in-

service database. Similar as BR-3 using Gibbs sampler, described in chapter 4, based on this data, the model weights of each unit, $\mathbf{w}_j$, are estimated using OLS method.

The parameters η and Λ^{-1} can be estimated by calculating the mean and covariance of all the $\mathbf{w}_j$ obtained by OLS, respectively. The parameter, ρ, is restricted to $\rho > D-1$ to ensure that the Gamma function in the normalisation factor is well-defined (appendix A.4). Because the mean of Wishart distribution of Λ, is $\Lambda = \rho\, R$ [Murphy, 2007], the prior parameter R can be computed from $R = \rho^{-1}\, \Lambda$.

In order to estimate prior parameters of equation (5.9), firstly it is necessary to take the squared residual values between predicted signal using OLS and the real measurement for the sample data. Next, mean, $\hat{\mu}_j$, and variance, $\hat{\sigma}_j^2$, of the squared residual values can be computed. Because the squared residual values are assumed to be IG distribution and using the relationship: $\beta \sim \text{Gam}(a, b)$ and $\beta = 1/\sigma^2$, so that $\sigma^2 \sim \text{IG}(a, b)$ [Murphy, 2007], therefore the parameters, a_j^β and b_j^β of equation (5.9), can be calculated by: $a_j = (\hat{\mu}_j/\hat{\sigma}_j)^2 + 2$ and $b_j = (\hat{\mu}_j^3/\hat{\sigma}_j^2) + \hat{\mu}_j$. For a new unit, these parameters can be initiated by computing the mean of training prior parameters, a_j^β and b_j^β.

The final prior parameters of a^α and b^α of equation (5.12) are estimated from the mean of a_j^β and b_j^β, respectively.

5.2.4 Likelihood function

Multiple observed degradation signals are used to compute the likelihood function. The likelihood for our model is conditional on the joint probability of observing the data ($\mathbf{X}_j$) and the model parameters ($\mathbf{w}_j, \sigma_j^2$), written:

$$p(\mathbf{y}_j|\mathbf{w}_j, \beta_j^{-1}, \mathbf{X}_j) = \mathcal{N}(\mathbf{y}_j|\mathbf{\Phi}_j\mathbf{w}_j, \beta_j^{-1}I) \tag{5.13}$$

where, j is the unit number.

5.2.5 Variational posterior distributions

A joint distribution of all the variables is given by:

$$p(\mathbf{y}_j, \mathbf{w}_j, \bar{\mathbf{w}}, \Lambda, \beta_j, \alpha) = \left[\prod_{j=1}^{M} p(\mathbf{y}_j|\mathbf{w}_j, \beta_j^{-1}, \mathbf{X}_j)\, p(\mathbf{w}_j|\bar{\mathbf{w}}, \beta_j, \Lambda)\, p(\beta_j|a_j^{\beta}, b_j^{\beta})\right] p(\bar{\mathbf{w}}|\eta, \alpha, \Lambda)\, p(\Lambda|\rho, R)\, p(\alpha|a^{\alpha}, b^{\alpha}) \tag{5.14}$$

where equation (5.14) can be represented as a directed graphical model as shown in figure 5.3.

For simplification, we define $\theta = \{\mathbf{w}_j, \bar{\mathbf{w}}, \Lambda, \beta_j, \alpha\}$. In order to find an approximation to the posterior, $q(\theta)$ is factorised according to equation (5.4), to give:

$$q(\theta) = q(\alpha)q(\bar{\mathbf{w}}, \Lambda)\prod_{j}\{q(\mathbf{w}_j), q(\beta_j)\} \tag{5.15}$$

which makes the assumption that $\mathbf{w}_j, \{\bar{\mathbf{w}}, \Lambda\}, \beta_j$ and α are independent in the posterior. According to the general results shown in equation (5.5), the logarithm of the joint distribution, $p(\mathbf{y}_j, \theta)$, equation (5.14) can be taken, and by using probability product rule, this can be decomposed as:

$$\begin{aligned} \ln p(\mathbf{y}_j, \theta) &= \ln\left[p(\mathbf{y}_j|\theta)\, p(\theta)\right] \\ &= \ln\left[p(\mathbf{y}_j|\theta)\right] + \ln\left[p(\theta)\right] \end{aligned} \tag{5.16}$$

and substituting from equations (5.8)-(5.13), gives:

$$\begin{aligned}\ln p(\mathbf{y}_j,\theta) = &-\frac{1}{2}\sum_{j=1}^{M}\{\beta_j(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)\}\\ &-\frac{1}{2}\sum_{j=1}^{M}\{\beta_j(\mathbf{w}_j-\bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})\}\\ &-\frac{1}{2}\alpha(\bar{\mathbf{w}}-\eta)^T\Lambda(\bar{\mathbf{w}}-\eta)+\frac{1}{2}(\rho+MD-1)\ln|\Lambda|\\ &-\frac{1}{2}\mathrm{tr}(R^{-1}\Lambda)\\ &+\sum_{j=1}^{M}[\{\frac{1}{2}(N_j+MD)+(a_j^{\beta}-1)\}\ln(\beta_j)-b_j^{\beta}\beta_j]\\ &+(a^{\alpha}-1+\frac{1}{2}D)\ln(\alpha)-b^{\alpha}\alpha \end{aligned} \tag{5.17}$$

where D, N_j, M are the dimensions of model coefficient's $\mathbf{w}_j$, the sample size for each unit and the total number of units, respectively.

Next, in order to obtain the optimal factors for the posterior approximation, equation (5.15), the equation (5.17) is substituted into equation (5.5), and the approximated posterior distributions are given by:

$$q^*(\mathbf{w}_j) = \mathcal{N}(\mathbf{w}_j|d_{\mathbf{w}_j},(\mathbb{E}[\beta_j])^{-1}D_{\mathbf{w}_j}) \tag{5.18}$$

$$q^*(\bar{\mathbf{w}}|\Lambda) = \mathcal{N}(\bar{\mathbf{w}}|d_{\bar{\mathbf{w}}},D_{\bar{\mathbf{w}}}) \tag{5.19}$$

$$q^*(\Lambda) = \mathcal{W}(\Lambda|\hat{\rho},\hat{R}) \tag{5.20}$$

$$q^*(\beta_j) = \mathrm{Gam}(\beta_j|\hat{a}_j^{\beta},\hat{b}_j^{\beta}) \tag{5.21}$$

$$q^*(\alpha) = \mathrm{Gam}(\alpha|\hat{a}^{\alpha},\hat{b}^{\alpha}) \tag{5.22}$$

where

$$D_{\mathbf{w}_j}^{-1} = (\mathbf{\Phi}_j^T \mathbf{\Phi}_j + \mathbb{E}[\Lambda]) \tag{5.23}$$

$$d_{\mathbf{w}_j} = D_{\mathbf{w}_j}(\mathbf{\Phi}_j^T \mathbf{y}_j + \mathbb{E}[\Lambda] d_{\bar{\mathbf{w}}}) \tag{5.24}$$

$$r = (\mathbb{E}[\alpha] + \sum_{j=1}^{M} \mathbb{E}[\beta_j])^{-1} \tag{5.25}$$

$$D_{\bar{\mathbf{w}}} = r\Lambda^{-1} \tag{5.26}$$

$$d_{\bar{\mathbf{w}}} = r(\sum_{j=1}^{M} \mathbb{E}[\beta_j] d_{\mathbf{w}_j} + \mathbb{E}[\alpha]\eta) \tag{5.27}$$

$$\hat{\rho} = \rho + M \tag{5.28}$$

$$\hat{R}^{-1} = R^{-1} + \sum_{j=1}^{M} (D_{\mathbf{w}_j} + \mathbb{E}[\beta_j] d_{\mathbf{w}_j} d_{\mathbf{w}_j}^T) + \mathbb{E}[\alpha]\eta\eta^T - r^{-1} d_{\bar{\mathbf{w}}} d_{\bar{\mathbf{w}}}^T \tag{5.29}$$

$$\hat{a}_j^{\beta} = \frac{N_j + MD}{2} + a_j^{\beta} \tag{5.30}$$

$$\begin{aligned} \hat{b}_j^{\beta} = b_j^{\beta} + \frac{1}{2} \sum_{j=1}^{M} \{ & (d_{\mathbf{w}_j} - d_{\bar{\mathbf{w}}})^T \mathbb{E}[\Lambda] (d_{\mathbf{w}_j} - d_{\bar{\mathbf{w}}}) \\ & + (\mathbf{y}_j - \mathbf{\Phi}_j d_{\mathbf{w}_j})^T (\mathbf{y}_j - \mathbf{\Phi}_j d_{\mathbf{w}_j}) \\ & + (\mathbb{E}[\beta_j])^{-1} \operatorname{tr}(\mathbf{\Phi}_j^T \mathbf{\Phi}_j D_{\mathbf{w}_j}) \\ & + (\mathbb{E}[\beta_j])^{-1} \operatorname{tr}(\Lambda D_{\mathbf{w}_j}) + rD \} \end{aligned} \tag{5.31}$$

$$\hat{a}^{\alpha} = a^{\alpha} + \frac{D}{2} \tag{5.32}$$

$$\hat{b}^{\alpha} = b^{\alpha} + \frac{1}{2} \{ (d_{\bar{\mathbf{w}}} - \eta)^T \mathbb{E}[\Lambda] (d_{\bar{\mathbf{w}}} - \eta) + rD \} \tag{5.33}$$

where $\mathbb{E}[\beta_j]$ and $(\mathbb{E}[\beta_j])^{-1}$ are the expectations and the reciprocal of the expectations of Gamma distribution of parameter β_j, respectively. $\mathbb{E}[\Lambda]$ and $\mathbb{E}[\alpha]$ are the expectations of Wishart and Gamma distributions of parameters Λ and α, respectively. The detailed derivation of the approximated posterior distributions is described in appendix D.2.

5.2.6 Variational Lower bound

The variational lower bound is useful to monitor the lower bound for our model, during re-estimation in order to test for convergence. By evaluating the equation (5.2), variational lower bound, $\mathcal{L}(q)$, can be computed from:

$$\begin{aligned}\mathcal{L}(q) = &\frac{1}{2}(MD - M + D)\mathbb{E}[\ln|\Lambda|] - \frac{1}{2}\text{tr}(R^{-1}\mathbb{E}[\Lambda]) \\ &- \hat{a}^{\alpha}\ln|\hat{b}^{\alpha}| - \sum_{j=1}^{M}\hat{a}_j^{\beta}\ln|\hat{b}_j^{\beta}| \\ &+ \ln\Gamma(\hat{a}^{\alpha}) + \sum_{j=1}^{M}\ln\Gamma(\hat{a}_j^{\beta}) \\ &+ \frac{1}{2}\sum_{j=1}^{M}\ln|D_{\mathbf{w}_j}| - \frac{1}{2}\sum_{j=1}^{M}\ln|\mathbb{E}[\beta_j]| + D(1+\ln(2\pi)) + \frac{1}{2}\ln|D_{\bar{\mathbf{w}}}| \\ &+ \frac{\hat{\rho}}{2}\ln|\hat{R}| + \frac{\hat{\rho}D}{2}(1+\ln|2|) + \frac{1}{4}D(D-1)\ln|\pi| \\ &+ \sum_{i=1}^{D}\ln\Gamma\left(\frac{\hat{\rho}+1-i}{2}\right) \end{aligned} \tag{5.34}$$

where the detailed of derivation of variational lower bound can be found in appendix D.4.

5.2.7 Predictive distribution

The posterior predictive density of the future of a degradation signal, $\mathbf{y}_j^*$, can be computed deterministically by marginalising the approximated posteriors $q(\mathbf{w}_j)$ and $q(\beta_j)$:

$$\begin{aligned} p(\mathbf{y}_j^*|\mathbf{X}_j^*) &= \int\int p(\mathbf{y}_j^*|\mathbf{X}_j^*, \mathbf{w}_j, \beta_j)p(\mathbf{w}_j)p(\beta_j)\, d\mathbf{w}_j d\beta_j \\ &= \int\int p(\mathbf{y}_j^*|\mathbf{X}_j^*, \mathbf{w}_j, \beta_j)q(\mathbf{w}_j)q(\beta_j)\, d\mathbf{w}_j d\beta_j \\ &= \int \mathcal{N}(\mathbf{\Phi}_j^* d_{\mathbf{w}_j}, \beta_j^{-1}(I + \mathbf{\Phi}_j^* D_{\mathbf{w}_j}\mathbf{\Phi}_j^{*T})\text{Gam}(\beta_j|\hat{a}_j^{\beta}, \hat{b}_j^{\beta})d\beta_j \end{aligned} \tag{5.35}$$

where $\mathbf{X}_j^*$ is future input value for j^{th} unit.

This results in a predictive Student-t distribution, given by:

$$p(\mathbf{y}_j^*|\mathbf{X}^*) = \text{St}\left(\mathbf{\Phi}_j^* d_{\mathbf{w}_j}, (\mathbb{E}[\beta_j])^{-1}(I + \mathbf{\Phi}_j^* D_{\mathbf{w}_j}\mathbf{\Phi}_j^{*T}), 2a_j^{\beta}\right) \tag{5.36}$$

where the properties of this distribution are:

$$\mathbb{E}[\mathbf{y}_j^*] = \mathbf{\Phi}_j^* d_{\mathbf{w}_j} \tag{5.37}$$

$$\text{cov}[\mathbf{y}_j^*] = \frac{b_j^\beta}{a_j^\beta - 1}(I + \mathbf{\Phi}_j^* D_{\mathbf{w}_j} \mathbf{\Phi}_j^{*T}) \tag{5.38}$$

$$\nu = 2a_j^\beta \tag{5.39}$$

where $\mathbb{E}[\mathbf{y}_j^*]$, $\text{cov}[\mathbf{y}_j^*]$ and ν are the mean, covariance and degrees-of-freedom of the predictive Student-t distribution, respectively. The derivation of predictive distribution is described in detail in appendix D.3

5.2.8 Failure-time distribution and RUL

In order to calculate failure-time distribution, the probability that the estimated failure time (T) exceeds time (t) is equivalent to the probability that on future degradation signal for unit j^{th} (y_j^*) is less than some failure threshold (y_{fail}). This can be formulated as:

$$P(T > t) = P(\text{y}_j^* \leq \text{y}_{\text{fail}}) \tag{5.40}$$

Therefore, the failure-time distribution can be written as:

$$\begin{aligned} F(t) &= P(T \leq t) = 1 - P(T > t) \\ &= 1 - P(\mathbf{y}_j^* \leq \text{y}_{\text{fail}}) \\ &= 1 - F_a\left(\frac{\text{y}_{\text{fail}} - \mathbf{\Phi}_j^* d_{\mathbf{w}_j}}{\sqrt{b_j^\beta (I + \mathbf{\Phi}_j^* D_{\mathbf{w}_j} \mathbf{\Phi}_j^{*T})}}, a_j^\beta\right) \end{aligned} \tag{5.41}$$

$$f(t) = \frac{dF(t)}{dt} \tag{5.42}$$

where $F_a(.)$ is the cumulative distribution function (cdf) of the Student-t distribution. Because $f(t)$ is derived from predictive Student-t distribution, it can be regarded as an extension of Beirnstein distribution, described in chapter 3.

The final step is estimating RUL by subtracting the current time and the mode of failure-time distribution $f(t)$, known as the estimated failure time, denoted by T. The current time is denoted by τ. Thus, the RUL is then given by: $\text{RUL} = T - \tau$.

5.3 Case studies

This section presents analysis of prognostic methods using BR-3 (with Gibbs sampler), described in chapter 4, and BR-3-VB, applied to synthetic and real TGT margin data. BR-2 algorithm, described in chapter 3, is also included in the comparison, to emphasise the strengths of BR-3 and BR-3-VB.

As mentioned in the previous studies, real TGT margin degradation data is difficult to use, directly, for comparing prognostic algorithms, because they are usually right-censored and so do not have ground truth for failure time. Therefore, synthetic degradation data of known properties is generated to test the algorithms. Subsequently, the algorithms are applied to the real TGT margin data for validation.

5.3.1 Case study 1: Synthetic degradation data

In this subsection, the same synthetic data as case study 1 in section 4.3.1 are used. As shown in figure 4.4, signals for units 2 and 11 are still selected for prognostics because they represent the extremes of behaviour in the signal. Unit 2 represents a noisy degradation signal with a sudden rate change in its behaviour whereas unit 11 has very high noise level.

The prior setting for BR-3-VB is obtained from description in section 5.2.3, whereas BR-2 and BR-3, the same prior specification are used as described in chapter 3 and 4, respectively. For BR-3, the iteration for Gibbs sampling is set to be $G_s = 10,000$ with burn-in time period is determined to be $B = 1,000$. The first-order polynomial basis function is still used for all methods.

Figure 5.4 shows the trajectory (degradation) prediction of synthetic data for unit 2 from time index 0.25 onward, i.e. inference is performed only on data for unit 2, up to time index 0.25. In this top subfigure, the straight dashed line, the straight solid line and the straight dot-dashed line represent the results of the BR-2, BR-3 and BR-3-VB, respectively. The symbol "+" represents the noisy health index. The action warning threshold ($y_{fail} = 22$) is illustrated by the horizontal dashed line and current time is shown by the vertical dashed line. In this lower subfigure, the dashed, solid, dot-dashed and dotted "bell-curves" represent the failure-time distribution of BR-2, BR-3, BR-3-VB and the ground truth, respectively.

This figure illustrates that the inference up to time index 0.25 is dominated by data with the shallow slope. It can be seen that BR-3 and BR-3-VB show similar performance. Both algorithms are able to maintain the prediction close to the true value because information sharing across units permitted by the additional hierarchical level, whereas BR-2 deviates strongly from the truth.

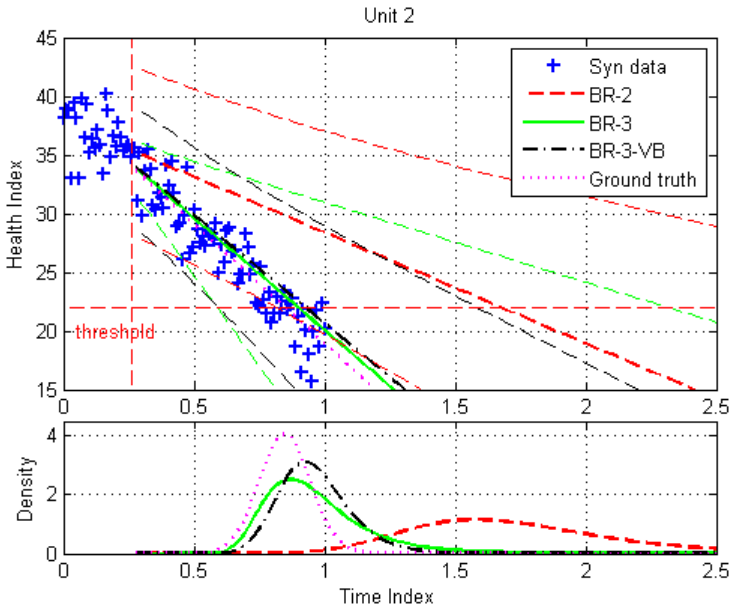


Figure 5.4: Trajectory (degradation) predictions of synthetic data for unit 2 at time index 0.25. The BR-3 and BR-3-VB predictions show similar behaviour.

Figure 5.5 illustrates the evolution of predicted failure-time distributions for unit 2 at time indexes 0.1, 0.73 and 1.49 for BR-2, BR-3 and BR-3-VB. The dotted "bell-curve" is the true failure-time distribution. It can be seen that the mode of failure-time distributions for BR-3 and BR-3-VB are similar where they do not deviate largely from true mode of failure-time distribution as time progresses. However, the mode of failure-time distribution for BR-2 deviates significantly in comparison to failure-time distributions of BR-3 and BR-3-VB at time index 0.73, but when enough data is gathered, BR-2 converges close to the truth.

Figure 5.6 demonstrates RUL metrics for units 2 and 11. The solid line is the true RUL and symbols "$\bigcirc$", "$+$" and "$\triangle$" represent RUL for BR-2, BR-3 and BR-3-VB, respectively. For both cases, BR-2 deviates strongly from the true value because it

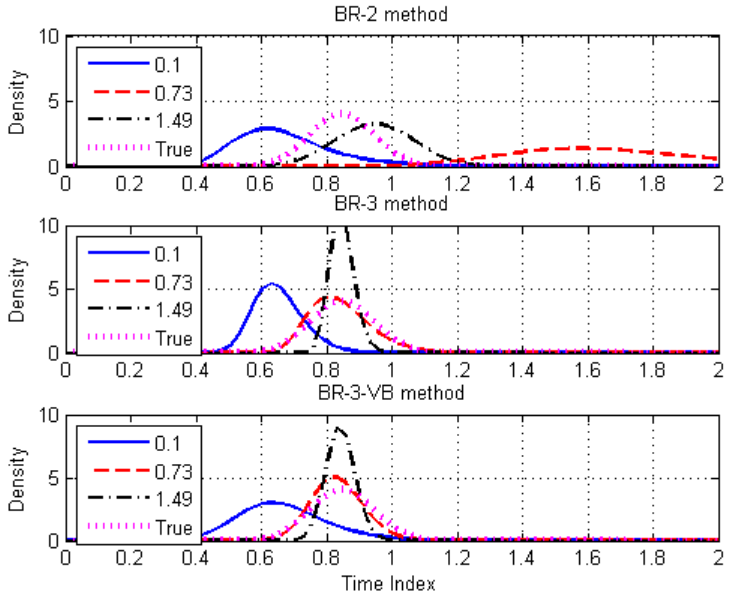
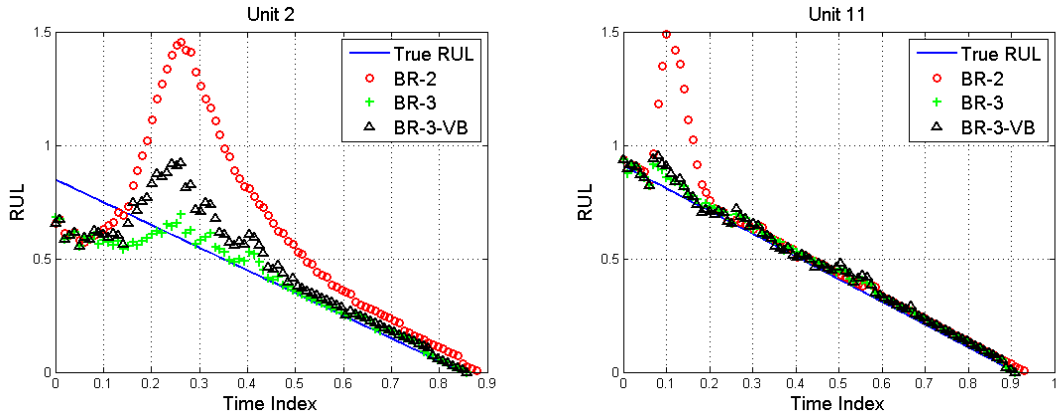


Figure 5.5: Failure-time distributions for BR-2, BR-3 and BR-3-VB. The mode of failure-time distributions of BR-3 and BR-3-VB are similar, indicate that the quality of approximation of BR-3-VB is satisfactory.

relies solely on individual prior. On the other hand, both BR-3 and BR-3-VB have similar performances, where they are able to maintain the RUL predictions close to the true RUL value owing to the existence of common prior component, which can share information across the units.

To summarise, performance across all units, the RUL residual is calculated for each individual at each time in turn. The RUL residual is the difference between estimated and true RUL. A comparison of RUL residuals for all data is illustrated by a time-indexed box plot in figure 5.7. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers. This figure emphasises that BR-3 and BR-3-VB have similar overall performance. The box plot for BR-2 shows more overall variability, especially in early stages (approximately before time index 0.4) compared to BR-3 and BR-3-VB. Some extreme scenarios, involved in degradation signals, lead to less accuracy and consistency in BR-2.

RUL difference between BR-3 and BR-3-VB is calculated for each unit. For all



(a) Unit 2 (b) Unit 11

Figure 5.6: RUL metrics for units 2 and 11 using BR-2, BR-3 and BR-3-VB. RUL for BR-3 and BR-3-VB show similar performance and they are able to maintain the RUL predictions close to the truth in comparison to RUL for BR-2.

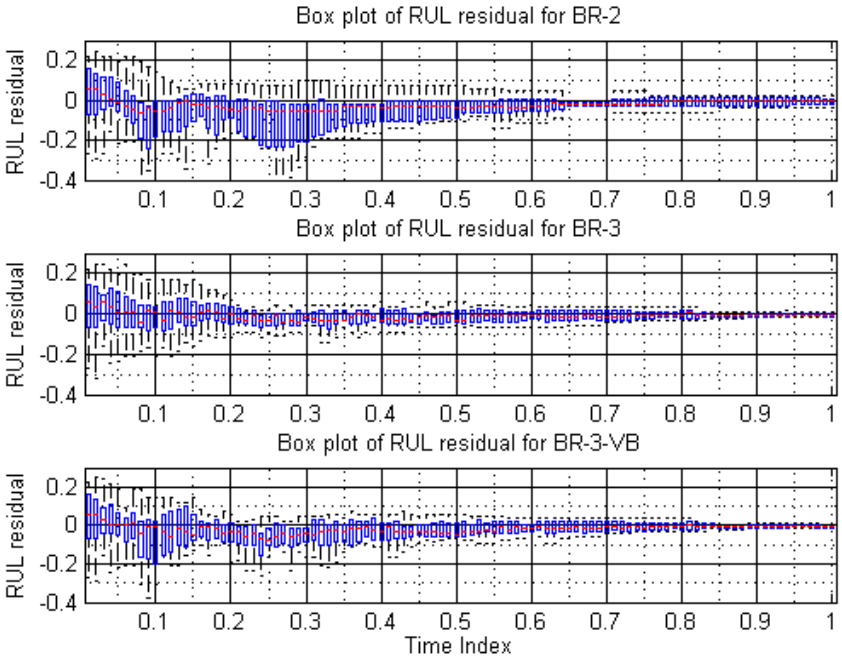


Figure 5.7: Box plots representing RUL performance for all synthetic data. BR-3 and BR-3-VB have similar overall performance. In general, BR-2 shows more overall variability, which indicates less consistency and accuracy in comparison to the BR-3 and BR-3-VB.

units, the average RUL difference is illustrated by a box plot, shown in figure 5.8. Both predictions start at almost the same level of zero point, due to the similarity in their prior distributions. Then, a slight deviation takes place because BR-3-VB cannot perform exactly as Gibbs sampler solution in BR-3, instead BR-3-VB approximates the posterior solutions. Ultimately, both approaches converge close to zero residual, which indicates that the variational solution is a satisfactory approximation to the full Bayesian inference.

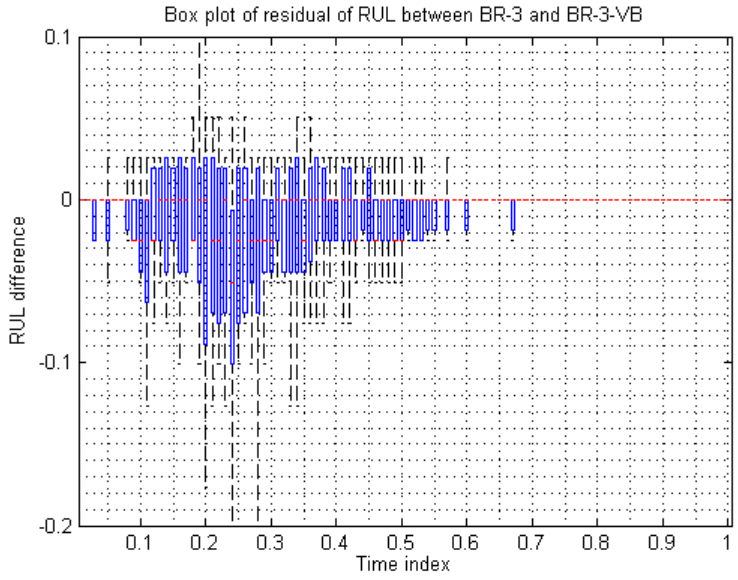
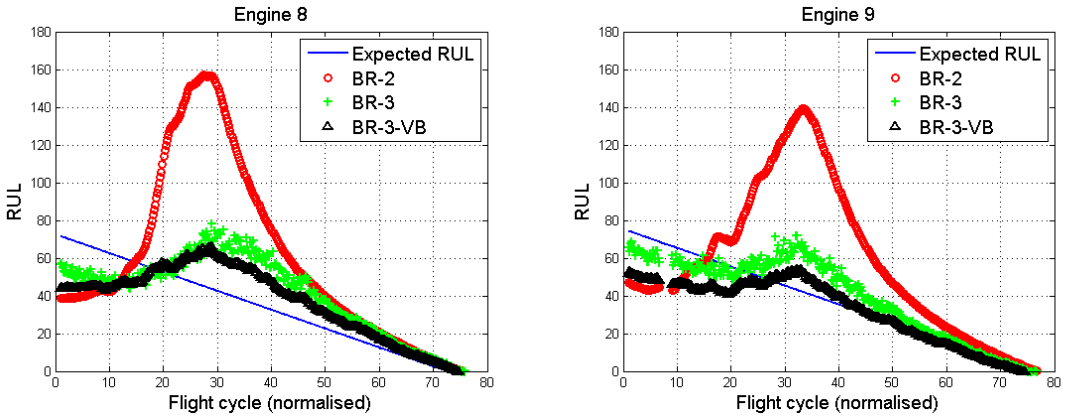


Figure 5.8: Box plot representing residual between RUL using BR-3 and BR-3-VB for all synthetic data. Owing to the similarity in their prior distributions, both predictions start at almost the same level. Next, they deviate because BR-3-VB cannot perform exactly as Gibbs sampler solution in BR-3, instead BR-3-VB approximates the posterior solutions. Ultimately, both approaches converge close to zero residual. This indicates that the variational solution is a satisfactory approximation to the full Bayesian inference.

5.3.2 Case study 2: Turbine Gas Temperature margin

This section focuses on real case study, prognostics based on operational or in-service data, the take-off values of TGT margin for an aerospace gas turbine engine. The key



(a) Engine 8 (b) Engine 9

Figure 5.9: RUL metrics for engines 8 and 9 using BR-2, BR-3 and BR-3-VB. BR-3 and BR-3-VB have similar performance in both RULs, indicate that the approximation quality of BR-3-VB is satisfactory.

objective is to illustrate the effectiveness of the proposed algorithm using real degradation data.

Here, TGT margin data, used in section 4.3.2 are still utilised. In particular, As shown in figure 4.11b, engines 8 and 9 are selected for prognostics because engine 8 has a sudden change in degradation behaviour, and engine 9, has very high noise level in the signal. This is what we attempted to deal with in the synthetic data.

Figure 5.9 shows the comparisons between BR-3 and BR-3-VB using RUL metric for the engines 8 and 9. The solid line is the "expected RUL", the symbols "$\bigcirc$", "$+$" and "$\triangle$" represent RUL for BR-2, BR-3 and BR-3-VB, respectively. It can be seen that BR-3 and BR-3-VB still demonstrate similar performance as in case of synthetic data, described in section 5.3.1. Figure 5.9a reveals that there is a significant deviation in performance of RUL for BR-2 due to the sudden change in degradation behaviour, whereas BR-3 and BR-3-VB can still maintain their predictions close to the "expected RUL" value. A similar result is also shown in figure 5.9b: the BR-2 prediction deviates strongly from the "expected RUL" owing to very high noise level in TGT margin data, whilst the BR-3 and BR-3-VB are still able to deal with this issue. After receiving additional data, RUL estimations for all methods converge to the "expected RUL" value.

Figure 5.10 shows box plots, representing RUL residual of BR-2, BR-3 and BR-3-

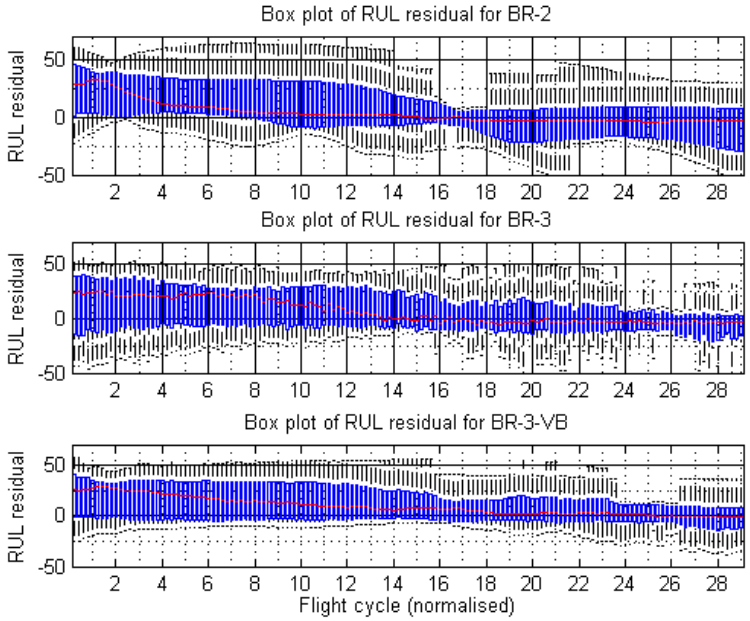


Figure 5.10: Box plots represents RUL performances for all real engine data. In general, the performance for BR-3 and BR-3-VB are better than BR-2 because they have less whisker and the edges of the box are smaller in comparison to box plot for BR-2.

VB when they are applied to all data available . In this figure, similar performance are still shown for BR-3 and BR-3-VB box plots. BR-2 has more variations than BR-3 and BR-3-VB (it can be seen from whisker). BR-3 and BR-3-VB have less whiskers and their edges of the box, representing the 25^{th} and 75^{th} percentiles, are smaller than BR-2. This concludes that, BR-3 and BR-3-VB have better overall performance. Figure 5.10 proves that the conclusions drawn from the synthetic data are borne out.

5.4 Summary

The main purpose of this chapter is to introduce a deterministic Bayesian hierarchical model, called BR-3 with variational Bayes (BR-3-VB). This is then compared to two existing inference approaches in Bayesian prognostic algorithms: BnHM using analytic solution (BR-2) and BHM using Gibbs sampler (BR-3), described in chapters 3 and 4, respectively. These approaches are implemented on synthetic degradation data to evaluate their relative performances. Subsequently, it is found that the characteristics,

observed in the synthetic environment, are preserved when applied to actual in-service data.

The BR-3 approach is appropriate to maximise the use of the health signal data generated by civil aerospace fleets. By modelling asset degradation with a hierarchical model, the health information can be shared between the engines in order to robustly estimate the RUL of the engines.

The two level hierarchical formulation BR-3 can be solved conventionally by Gibbs sampler. However, this method may not be attractive for use in the prognostic field owing to high computational cost and the level of expertise needed for determining some crucial parameters. To overcome these issues, BR-3-VB, that is BR-3 using variational inference, is developed, which delivers an approximate but deterministic solution. The results show that BR-3 and BR-3-VB have similar performance in RUL estimation. This approximation proves its capability in preserving prediction performance, reducing computational speed and eliminating expert involvement. Hence, variational Bayesian inference is a promising alternative to Gibbs sampling for an analytically intractable Bayesian model.

In conclusion, this chapter delivers three main contributions. Firstly, approximate solutions for Bayesian hierarchical model based on variational inference are derived. Secondly, Bayesian prognostic algorithms are applied for estimating RUL of aerospace gas turbine engine. The results for both prognostic algorithms based on BHM outperform BnHM. The final contribution is that variational Bayes demonstrates successfully its capability to substitute Gibbs sampler for BHM prognostics.

Chapter 6

Accommodating Irregular Events into Prognostics: A Bayesian framework

6.1 Introduction

Chapters 3-5 have addressed several challenges associated with a class of complex systems in particular gas turbine engine, described in section 1.2.3. First, data-driven method is a suitable prognostic approach for a complex system (section 1.2.3.1). Due to the absent of ground truth in real data (section 1.2.3.2), synthetic data of known properties are generated to emulate important characteristics of the real data for testing the developed prognostic techniques. Subsequently, the algorithms are applied to the real degradation data for validation.

In particular, Bayesian framework is a specific method of probabilistic approach which is able to cope with data uncertainty (section 1.2.3.3) where in previous works (chapters 3-5), various Bayesian techniques have demonstrated successfully their capability in dealing with uncertainty in degradation data.

Furthermore, to overcome the challenges of accommodating heterogeneous fleet data (section 1.2.3.4), the value of gathered data has been maximised by using a Bayesian hierarchical modelling structure that is capable of accommodating multiple degradation signals, gathered from multiple engines in a particular fleet. In this way, the health information can be shared across the engines for RUL estimation of a specific engine to cope with extreme degradation scenarios, including very high noise problem (section

1.2.3.3) as well as slope change in degradation (section 1.2.3.5).

Most of the challenges in gas turbine engine prognostics, described in sections 1.2.3.5 and 1.2.3.6, can be summarised into two points:

(i) Handling rapid change in degradation
A gas turbine engine may degrade at approximately constant rate lasting for a period of time followed by an increase in rate owing to operating condition, fault and a step change in covariate [Ackert, 2010]. Despite this issue has been addressed in chapters 4 and 5 through the use of BR-3 or BR-3-VB, detecting this event will be beneficial to enhance RUL estimation by reconfiguring prognostic model.

(ii) Coping with recoverable system
A recovery in the degradation may occur owing to maintenance action. However, airlines, which carry out maintenance activities, do not always provide maintenance information to an organisation that performs fleet management. This results in a difficulty to reset the prognostic algorithm. An automatic alert of this change should be made by detecting the change in degradation data [Skaf et al., 2013].

This chapter introduces an integration concept, combining Bayesian prognostic algorithm and event detection to solve wide class of prognostic issues as described above.

6.2 Integrated prognostic approach

This section describes the generic methodology of an integrated approach and demonstrates its capability using synthetic data. This will be validated in real case study through real world data in the next section.

6.2.1 Generic methodology

Several events may affect health index/degradation of complex systems. This effect can be described through two main explanations. First situation is when health index recovers due to maintenance action. Another situation occurs when the slope in health index changes which takes place owing to fault or step change in covariates. For example, if

a fault[1] has occurred in an engine, the engine performance may deteriorate faster than non-faulty engine.

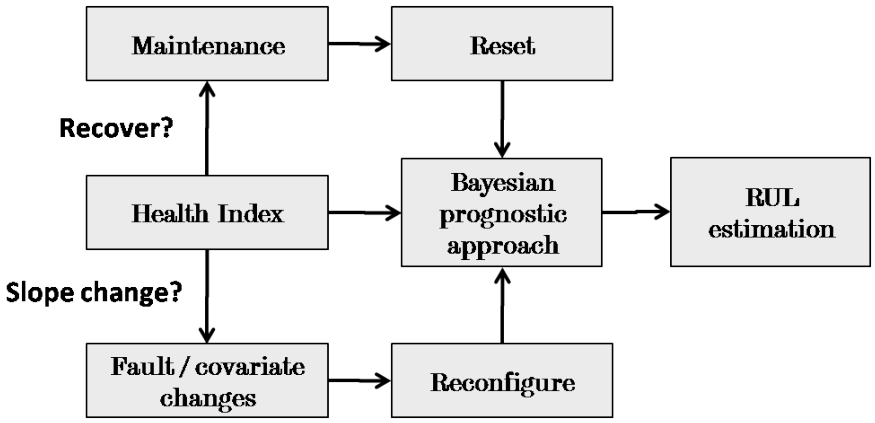


Figure 6.1: An illustration of generic prognostic block diagram. The knowledge about variation in health index is important to provide better RUL estimation through Bayesian prognostics.

A generic methodology about this concept is illustrated in figure 6.1. Prognostic algorithm can be reset if a major maintenance, e.g. compressor washing, has been carried out which results in an increase of health index. In this case, degradation database and prior parameters can be updated to then restart the RUL estimation.

In addition, when there is a slope change in health index, this information should be delivered to a main prognostic algorithm to provide better knowledge about how degradation behaviour would occur in the future. Therefore, the prognostic model should be reconfigured to improve RUL estimation.

For reconfiguration process, two degradation models are set up where each degradation model uses different type of prior parameters. But this concept is generic in nature as it can accommodate multiple degradation models based on the numbers of events detected in the degradation data. The first degradation model represents the system deterioration, which occurs at the normal rate, where the system degrades naturally due to ageing and wear. The prior parameters can be estimated through complete degradation cycle for all units. The second degradation model embodies the system deterioration, which occurs at the faster rate than the first model. In this case, the prior parameters are

[1]Here, this fault does not mean a catastrophic failure where there is a sudden and total failure in the system from which recovery is impossible.

calculated from the part of all degradation data available after the detected event.

Firstly, when the prognostic algorithm is started for a new unit[1], the first prior is utilised because the system is assumed to deteriorate in normal rate. Once the change is detected for a particular unit, the prognostic model of that unit will reconfigure its prior based on the second degradation model, where it contains the knowledge about the degradation behaviour after an event has occurred. This method is able to accommodate the information about any event occurring in any individual unit. Therefore, this should improve RUL estimation as we will see through a demonstration in this following section.

6.2.2 Demonstration 1: Synthetic degradation data

This subsection presents how the idea of integrated prognostic approach is able to solve two main issues described above. As in previous chapters, prior to validating using real data, synthetic degradation data of known properties are generated to test our proposed concept. Two methods are compared to demonstrate the efficacy of the proposed method. First method is BR-3-VB without any knowledge about the change in covariate or health index, whereas another method uses BR-3-VB combined with the knowledge of any changes in the degradation, named integrated BR-3-VB (Int-BR-3-VB).

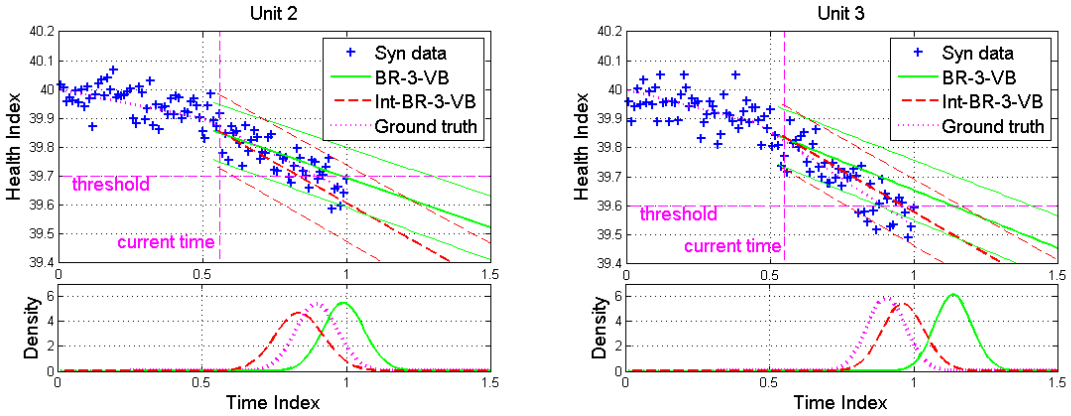
Two types of degradation scenario are generated. First scenario occurs when there is a slope change in degradation. Another scenario takes place when the degradation signal may recover due to maintenance event. In this case study, anomaly in covariate (fault or other step changes) as well as maintenance event are assumed to be known.

6.2.2.1 Capturing rapid change in degradation slope

In this case, the aim is to demonstrate how the quality of RUL estimation can be enhanced by reconfiguring prognostic algorithm if the step change in covariates or fault event is known. Synthetic data are generated from a piecewise linear model as described in equation (4.29). Here, various rates of change in degradation slope and time index where the change in degradation's slope occurs, x_c, can be modified.

Ten degradation data are generated with various x_c and rates of changes in degradation slope. Two units are selected, that are units 2 and 3, represent normal and rapid

[1] A new unit means a brand new unit or just maintained unit.



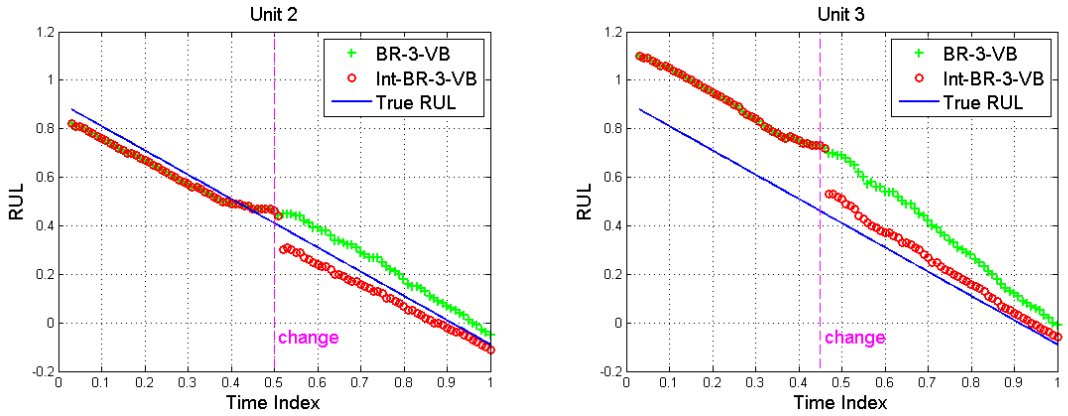
(a) Unit 2 (b) Unit 3

Figure 6.2: Trajectory predictions of BR-3 and Int-BR-3. The left figure is a degradation with slow decay and the right figure is degradation with rapid decay. Int-BR-3-VB can deal with rapid decay problem, because it adds extra knowledge about the change into the main prognostic algorithm.

change in degradation, respectively. All the setting for BR-3-VB, including prior parameters and basis function choice, is the same as in chapter 5.

Figure 6.2 shows two trajectory predictions for units 2 and 3. On the top subfigure, symbol "+" is the synthetic degradation data, whereas the solid line, dashed line and dotted line represent BR-3-VB, Int-BR-3-VB and ground truth, respectively. In the lower subfigure, solid bell-curve, dashed bell-curve and dotted bell-curve embody the failure-time distributions of BR-3-VB, Int-BR-3-VB and truth, respectively. It can be seen that when there is a slow decay (figure 6.2a), BR-3-VB demonstrates similar performance with Int-BR-3-VB. This is because BR-3-VB itself already has a common prior which is able to share the information across the engines to provides appropriate prediction for this scenario as described in chapters 4 and 5. In figure 6.2b however, BR-3-VB deviates largely from the truth whereas Int-BR-3-VB is able to retain the prediction close to the truth, because the second method adds extra knowledge about the change in health index into the main prognostic algorithm.

Figure 6.3 shows two RUL metrics for both scenarios. Solid line is the true RUL, whereas symbols "+" and "○" represent RUL estimations for BR-3-VB and Int-BR-3-VB, respectively. The vertical dashed line indicates where the slope change has oc-



(a) Unit 2

(b) Unit 3

Figure 6.3: RUL metrics for BR-3 and Int-BR-3 for two different scenarios. For rapid decay problem (right), Int-BR-3-VB converges faster than BR-3-VB, because Int-BR-3-VB includes additional knowledge about the change into the main prognostic algorithm.

curred. It can be seen that for slow decay problem (figure 6.3a), there is no substantial different between RUL for BR-3-VB and Int-BR-3-VB whilst for the rapid decay problem (figure 6.3b), RUL for Int-BR-3-VB converges faster than BR-3-VB, because once the change is known, the prognostic model in Int-BR-3-VB is reconfigured based on the second degradation model, which contains the knowledge about degradation behaviour after a detected event.

As in previous chapters, in order to prove the consistency of this integration concept, residuals of RUL for 10 different units are computed. Figure 6.4 shows the representation of RUL residual as box plot. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers. The top subfigure is a box plot of RUL residual for BR-3-VB, whilst the lower subfigure is a box plot for Int-BR-3-VB. In general, it can be seen that once the change occurs, approximately at time index 0.3, BR-3-VB has larger edges which indicate there is a lack of consistency in RUL residual in dealing with the slope changes in degradation signal, whereas in Int-BR-3-VB, the edges of box plot shows less variation which indicate that this method is more consistent to deal with slow and rapid decay problems.

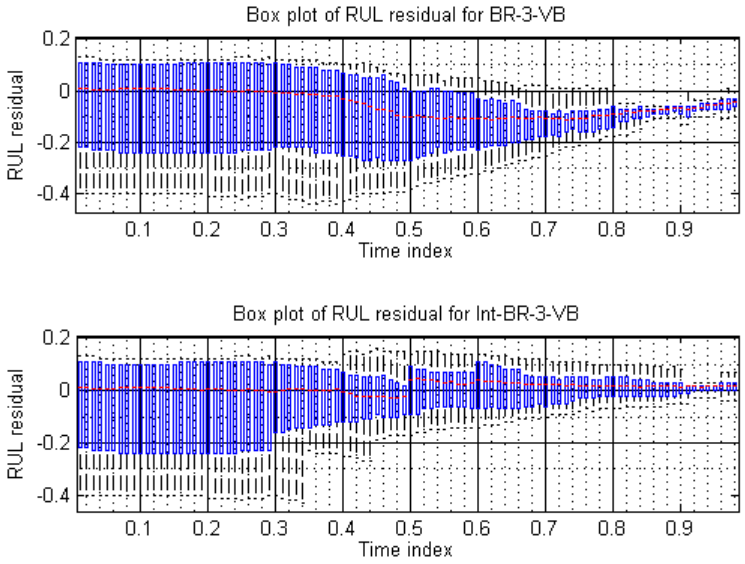


Figure 6.4: Box plots representing comparison of overall performance between BR-3-VB and Int-BR-3-VB. Int-BR-3-VB has better performance, indicated by smaller edges, which means that this method is able to deal with slow and rapid decay in degradation.

6.2.2.2 Handling maintenance event in a recoverable system

Prognostic algorithm requires to know maintenance event to reset RUL estimation. Here, the aim is to demonstrate how prognostic algorithm can be reset if the maintenance event is known.

Ten degradation data is generated using equation (3.18), but a degradation cycle (unit 10) is added into unit 1 to demonstrate that this unit can be recovered. Therefore unit 1 comprises two degradation cycles, where it recovers at time index one. This unit is shown in figure 6.5, where symbol "+" is the noisy degradation data and the solid line represents its ground truth.

If the knowledge about when the maintenance event is known, the prior distribution parameters are reset and therefore RUL can be re-estimated. Figure 6.6 shows RUL metric for unit 1. Solid line and symbol "○" represent true RUL and RUL for BR-3-VB, respectively. At time index one (once the health index recovers), the first degradation cycle is added into degradation database and then prior parameters are re-estimated. It can be seen that the new initial prediction starts in approximated prediction based on

updated prior parameters, and then it converges close to the true RUL.

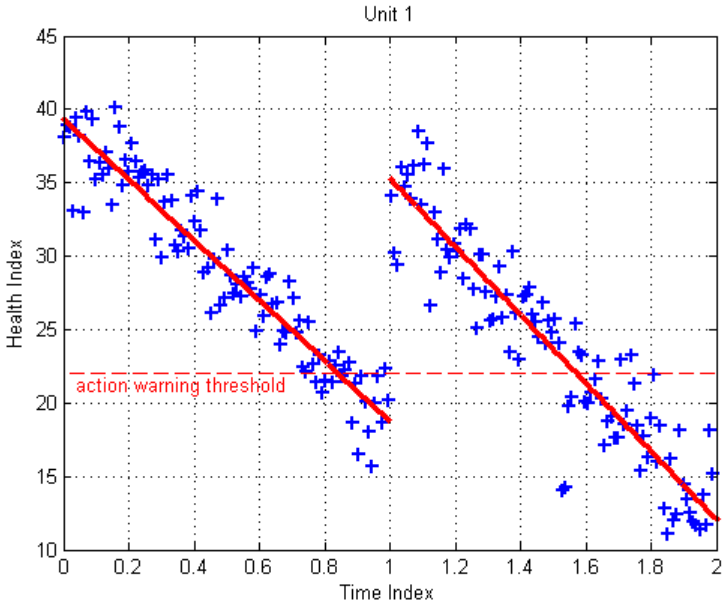


Figure 6.5: An example of synthetic degradation signal where it recovers at time index 1. The symbol "+" is the health index and the solid line is the ground truth.

6.3 Real case study

This section describes the implementation of the proposed concept in practice. The main challenge in this implementation is in the difficulty of obtaining information about the events which affect degradation behaviour.

The first challenge is as described in section 1.2.3.6, maintenance event may not be known in practice, where the uncertainty surrounding maintenance actions impacts the ability to determine starting and stopping points in prognostic analysis [Brasco et al., 2013].

Furthermore, it is difficult to detect any slope change directly in health indexes because they are typically very noisy. Another event at different level of system may also influence the slope change in health index, detecting this particular event at that level would provide better knowledge about how degradation behaviour of particular system would occur in the future.

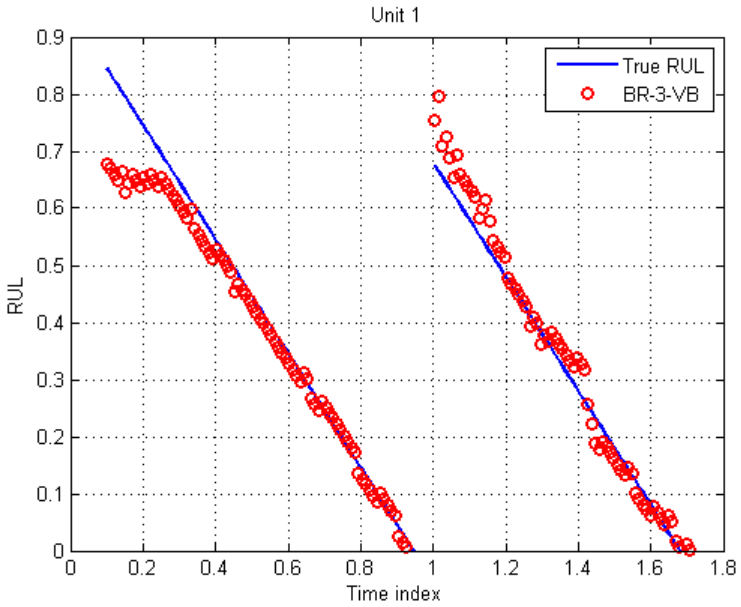


Figure 6.6: An example of synthetic degradation signal where it recovers at time index 1. The solid line is the true RUL whereas the symbol "○" is the RUL estimation for BR-3-VB.

The proposed concept uses, optimally, data available from multiple fleets of engine (figure 1.7) as well as from different levels of hierarchy in gas turbine engine architecture (figure 1.6) for estimating the RUL of a specific engine. In this work, BR-3-VB is still used, as a main prognostic algorithm because it is based on deterministic Bayesian hierarchical model with several advantages for complex system prognostics, described in chapter 5. In particular, a change point detection (CPD) algorithm is selected as event detection method, to detect any anomaly in time-series data.

Figure 6.7 shows a detailed block diagram of the proposed concept. Several monitored data, e.g. health index and other covariates, are fed into main prognostic algorithm to estimate RUL of gas turbine engine. At the same time, these parameters are also monitored continuously by CPD algorithm. If CPD algorithm detects significant increase in TGT margin (the indicator score is bigger than the defined threshold), it considers that maintenance action is just performed and prognostic algorithm should be restarted.

Furthermore, whenever the CPD algorithm detects abnormality in one of monitored covariates (the indicator score is bigger than the defined threshold), it considers that

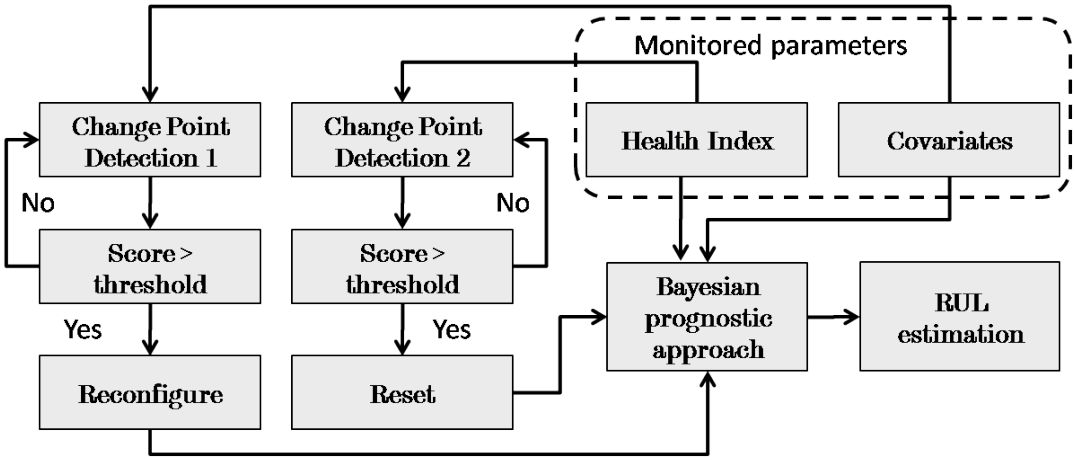


Figure 6.7: Block diagram of combining between Bayesian prognostic approach and CPD algorithm.

there is a fault or other step changes in operating conditions, which would affect the degradation. Therefore, the prognostic model should be reconfigured to improve the RUL estimation, as described in section 6.2.1.

Firstly below, the CPD algorithm used in this proposed approach is described thoroughly and after that a real case study is presented.

6.3.1 Change point detection by relative density-ratio estimation

A change point detection (CPD) algorithm, aims to discover points at which sudden changes occur in time-series data [Kawahara and Sugiyama, 2012; Liu et al., 2013]. This method can be classified based on the delay in detection: real-time detection or retrospective detection. Real-time detection is used for applications which require immediate response. On the other hand, retrospective detection can be used for applications, which tolerate longer reaction periods. The latter algorithm tends to give more robust and accurate detection [Liu et al., 2013].

For the case of gas turbine engine prognostics, where the delay has minor effect on the decision making, a retrospective detection algorithm, therefore, is suitable to monitor and detect the changes in degradation, e.g. TGT margin, as well as its covariates. In this application, accurate detection is crucial and the prediction is also updated every flight cycle, providing more than adequate time to compute the new prediction.

In this work, a CPD method, named relative unconstrained least-squares importance fitting (RuLSIF) [Liu et al., 2013; Yamada et al., 2013], is used to detect anomalies (maintenance and step change/fault events), mentioned earlier. This approach does not estimate probability densities, such as kernel density estimation [Brodsky and Darkhovsky, 1993; Csörgő and Horvath, 1988], but instead estimates the ratio of probability densities directly [Vapnik, 1998]. The following subsections summarise a CPD algorithm, based on RuLSIF [Yamada et al., 2013].

6.3.1.1 Relative Pearson divergence

Let assume $\mathbf{z}_t$ is a time-series of data to be monitored for detecting changes in the statistical properties of the data. In gas turbine engine's case, the observed data can be TGT margin or other relevant covariates. If the observed data for CPD algorithm is TGT margin, then $\mathbf{z}_t$ is the same as $\mathbf{y}$, shown in equation (3.1).

Parameter $\mathbf{z}_t$ is d-dimensional time-series sample at time t, where $\mathbf{z}_t = [z_1 \ z_2 \cdots z_d]$. Let assume $\mathbf{Z}_t$ as a sample of time series at time t with length k, given by:

$$\mathbf{Z}_t = [\mathbf{z}_t^T \ \mathbf{z}_{t+1}^T \cdots \mathbf{z}_{t+k-1}^T]^T \tag{6.1}$$

Next, let $\mathfrak{Z}_t$ be a set of n retrospective subsequence samples starting at time t:

$$\mathfrak{Z}_t = \{\mathbf{Z}_t \ \mathbf{Z}_{t+1} \cdots \mathbf{Z}_{t+n-1}\} \tag{6.2}$$

As illustrated in figure 6.8, this forms a $dk \times n$ *Hankel* matrix and this plays a key role in CPD [Kawahara et al., 2007; Moskvina and Zhigljavsky, 2003].

The methodology is to calculate a certain dissimilarity measure between two consecutive segments, $\mathfrak{Z}_t$ and $\mathfrak{Z}_{t+n}$, to be a basis of change points. The higher the dissimilarity measure the more likely it is that the change point has occurred. In RuLSIF, the dissimilarity measure is defined by:

$$D(P_t||P_{t+n}) + D(P_{t+n}||P_t) \tag{6.3}$$

where P_t and P_{t+n}, denoted by $P(\mathbf{Z})$ and $P_\alpha^{'}(\mathbf{Z})$, are probability distributions of samples in $\mathfrak{Z}_t$ and $\mathfrak{Z}_{t+n}$, respectively. $D(P_t||P_{t+n})$ is $\alpha-$relative Pearson (PE) divergence

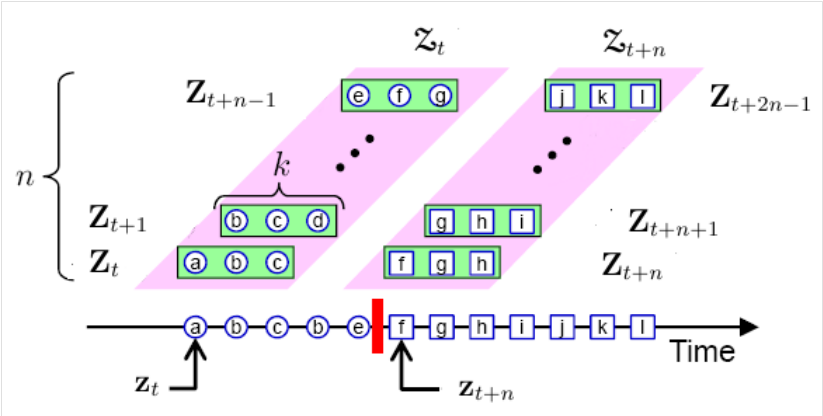


Figure 6.8: An illustration of example of notations on one-dimensional time-series data [Liu et al., 2013].

[Yamada et al., 2013], defined by:

$$D(P_t||P_{t+n}) = \text{PE}_\alpha(P||P') = \frac{1}{2}\,\text{PE}(P||\alpha\,P + (1-\alpha)\,P') \\ = \frac{1}{2}\int P_{\alpha'}(\mathbf{Z})\left(\frac{P(\mathbf{Z})}{P_{\alpha'}(\mathbf{Z})} - 1\right)^2 d\mathbf{Z} \tag{6.4}$$

where $\alpha-$relative PE divergence measures the difference between two probability distributions $P(\mathbf{Z})$ and $P_{\alpha'}(\mathbf{Z})$ for $0 \leq \alpha < 1$. The $\alpha-$relative density ratio, r_α, is defined by:

$$r_\alpha(\mathbf{Z}) = \frac{P(\mathbf{Z})}{\alpha P(\mathbf{Z}) + (1-\alpha)P'(\mathbf{Z})} = \frac{P(\mathbf{Z})}{P_{\alpha'}(\mathbf{Z})} \tag{6.5}$$

where $P_{\alpha'}(\mathbf{Z}) = \alpha P(\mathbf{Z}) + (1-\alpha)P'(\mathbf{Z})$ is the $\alpha-$mixture density and the α-relative density ratio is bounded above by $1/\alpha$ for $\alpha > 0$. The expectation of $f(\mathbf{Z})$ under $P(\mathbf{Z})$ is denoted by $\mathbb{E}_{P(\mathbf{Z})}[f(\mathbf{Z})]$, given by:

$$\mathbb{E}_{P(\mathbf{Z})}[f(\mathbf{Z})] = \int f(\mathbf{Z})P(\mathbf{Z})d\mathbf{Z} \tag{6.6}$$

Therefore, $\alpha-$relative PE divergence, equation (6.4), results in:

$$D(P_t||P_{t+n}) = \text{PE}_\alpha(P||P') = \frac{1}{2}\int P_{\alpha'}(\mathbf{Z})\big(r_\alpha(\mathbf{Z}) - 1\big)^2\, d\mathbf{Z} \\ = \frac{1}{2}\,\mathbb{E}_{P_{\alpha'}(\mathbf{Z})}\big[\big(r_\alpha(\mathbf{Z}) - 1\big)^2\big] \tag{6.7}$$

with this formulation, only the ratio need be calculated.

6.3.1.2 Density ratio model

The $\alpha-$relative density ratio, $r_\alpha(\mathbf{Z})$, can be modelled by the density ratio, $\frac{P(\mathbf{Z})}{P'(\mathbf{Z})}$, as a kernel, defined by:

$$r_\alpha(\mathbf{Z}) = g(\mathbf{Z}) = \sum_{l=1}^{n} \theta_l K(\mathbf{Z}, \mathbf{Z}_l) \tag{6.8}$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)^T$ are learning parameters and $K(\mathbf{Z}, \mathbf{Z}')$ is a Gaussian kernel, defined by:

$$K(\mathbf{Z}, \mathbf{Z}') = \exp\big(-\frac{||\mathbf{Z} - \mathbf{Z}'||^2}{2\gamma^2}\big) \tag{6.9}$$

where $\gamma(>0)$ is the kernel width, which is determined based on cross validation.

6.3.1.3 Learning algorithm

The parameters $\boldsymbol{\theta}$ can be learned by minimising the squared loss between true relative ratio, $r_\alpha(\mathbf{Z})$, and estimated relative ratio, $\hat{g}(\mathbf{Z})$, given by:

$$\begin{aligned} J(\mathbf{Z}) &= \frac{1}{2}\int P_{\alpha'}(\mathbf{Z})\big(r_\alpha(\mathbf{Z}) - \hat{g}(\mathbf{Z})\big)^2 d\mathbf{Z} \\ &= \frac{1}{2}\int P_{\alpha'}(\mathbf{Z}) r_\alpha(\mathbf{Z})^2 d\mathbf{Z} - \int P_{\alpha'}(\mathbf{Z}) r_\alpha(\mathbf{Z}) \hat{g}(\mathbf{Z}) d\mathbf{Z} + \frac{1}{2}\int P_{\alpha'}(\mathbf{Z}) \hat{g}(\mathbf{Z})^2 d\mathbf{Z} \\ &= \frac{1}{2}\int P_{\alpha'}(\mathbf{Z}) r_\alpha(\mathbf{Z})^2 d\mathbf{Z} - \int P(\mathbf{Z}) \hat{g}(\mathbf{Z}) d\mathbf{Z} + \frac{\alpha}{2}\int P(\mathbf{Z}) \hat{g}(\mathbf{Z})^2 d\mathbf{Z} \\ &\quad + \frac{1-\alpha}{2}\int P'(\mathbf{Z}) \hat{g}(\mathbf{Z})^2 d\mathbf{Z} \end{aligned} \tag{6.10}$$

First term is a constant because it is unrelated to $\hat{g}(\mathbf{Z})$, hence:

$$J(\mathbf{Z}) = -\mathbb{E}_{P(\mathbf{Z})}[\hat{g}(\mathbf{Z})] + \frac{\alpha}{2}\mathbb{E}_{P(\mathbf{Z})}[\hat{g}(\mathbf{Z})^2] + \frac{1-\alpha}{2}\mathbb{E}_{P(\mathbf{Z})}[\hat{g}(\mathbf{Z})^2] + \text{const} \tag{6.11}$$

By replacing $\hat{g}(\mathbf{Z})$, by a kernel, in equation (6.8), and approximating the expectation by empirical averages, then RuLSIF optimisation problem is given by:

$$\min_{\boldsymbol{\theta}\in R^n}\big[\frac{1}{2}\boldsymbol{\theta}^T\hat{\mathbf{H}}\boldsymbol{\theta} - \hat{\boldsymbol{h}}^T\boldsymbol{\theta} + \frac{\lambda}{2}\boldsymbol{\theta}^T\boldsymbol{\theta}\big] \tag{6.12}$$

where $\frac{\lambda}{2}\boldsymbol{\theta}^T\boldsymbol{\theta}$ is a penalty term for regulation purpose, $\lambda\,(\geq 0)$ is a regulation parameter. Parameters $\hat{\mathbf{h}}$ is is the n-dimensional vector with the l-th element given by:

$$\hat{h}_l = \frac{1}{n}\sum_{i=1}^{n} K(\mathbf{Z}_i, \mathbf{Z}_l) \tag{6.13}$$

and $\hat{\mathbf{H}}$ is the $n \times n$ matrix with the (l, l')-th element given by

$$\hat{H}_{l,l'} = \frac{\alpha}{n}\sum_{i=1}^{n} K(\mathbf{Z}_i, \mathbf{Z}_l)\, K(\mathbf{Z}_i, \mathbf{Z}_{l'}) \; + \frac{(1-\alpha)}{n}\sum_{j=1}^{n} K(\mathbf{Z}'_j, \mathbf{Z}_l)\, K(\mathbf{Z}'_j, \mathbf{Z}'_l) \tag{6.14}$$

Thus, the analytic solution of equation (6.12), can be found:

$$\hat{\boldsymbol{\theta}} = (\hat{\mathbf{H}} + \lambda I_n)^{-1}\hat{\boldsymbol{h}} \tag{6.15}$$

where I_n is the n-dimensional identity matrix. Finally, a density ratio estimator can be written as:

$$\hat{r}_\alpha(\mathbf{Z}) = \hat{g}(\mathbf{Z}) = \sum_{l=1}^{n} \hat{\theta}_l\, K(\mathbf{Z}, \mathbf{Z}_l) \tag{6.16}$$

when $\alpha = 0$, this method is reduced to unconstrained least-squares importance fitting (uLSIF) [Kanamori et al., 2009].

6.3.1.4 Change point detection by RuLSIF

In order to estimate $\alpha-$relative PE divergence, a density ratio estimator, equation (6.16), is substituted into PE-divergence, equation (6.7), which results in:

$$\hat{\text{PE}}_\alpha = -\frac{\alpha}{2n}\sum_{i=1}^{n}\hat{g}(\mathbf{Z}_i)^2 - \frac{(1-\alpha)}{2n}\sum_{j=1}^{n'}\hat{g}(\mathbf{Z}'_j)^2 + \frac{1}{n}\sum_{i=1}^{n}\hat{g}(\mathbf{Z}_i) - \frac{1}{2} \tag{6.17}$$

In order to detect the change-point, the symmetrised divergence is used, described in equation (6.3). Hence, the $\text{PE}_\alpha(\text{Symmetric})$ is defined as:

$$\text{PE}_\alpha(\text{Symmetric}) : \text{PE}_\alpha(P_t||P_{t+n}) + \text{PE}_\alpha(P_{t+n}||P_t) \tag{6.18}$$

In this case, $\text{PE}_\alpha(\text{Symmetric})$, equation (6.18), will be used as an indicator score, as a basis of change point score in TGT margin and its covariates to support our main prognostic algorithm, BR-3-VB.

6.3.1.5 Advantages

In summary, there are several reasons why this algorithm is suitable for detecting anomaly in gas turbine engine problems, which are:

(i) This approach is simple [Vapnik, 1998], because it does not estimate through density estimation, such as kernel density estimation, but this estimates the ratio of probability densities directly without going through density estimation[1].

(ii) The observed data (TGT margin and its covariates) contain large uncertainty. This method is based on non-parametric method, where it does not need to assume a specific functional form for the distribution. Therefore, such models can deal with model's complexity [Bishop, 2006; Rasmussen and Williams, 2006].

(iii) The solution can be computed analytically [Kanamori et al., 2009; Liu et al., 2013; Yamada et al., 2013].

(iv) The basic idea of RuLSIF is to consider relative density ratios, which are smoother and always bounded from above [Yamada et al., 2013]. Thresholds need to be defined to determine maintenance event as well as covariate changes. Therefore density ratios must be bounded.

[1]The rational behind the principle of direct density-ratio estimation is that knowing two densities means knowing their ratio, but not vice versa; knowing the ratio does not necessarily mean knowing the two densities [Liu et al., 2013].

6.3.2 Demonstration 2: Turbine Gas Temperature margin

In this part, the proposed concept, a combination of BR-3-VB and CPD will be validated directly through real world data. As demonstration using synthetic data, two prognostic scenarios are discussed in the following subsections.

6.3.2.1 Capturing rapid change in degradation slope

As described in section 1.2.3.5, a considerable change in degradation slope sometimes occurs in gas turbine engine, affected by various factors, including operating conditions, a step change in covariates or faults. There are several key operational parameters (covariates) used to monitor the performance of an engine, including engine's operating speed, temperature, pressure, fuel flow and vibration levels [Ackert, 2010].

In this work, the aforementioned CPD algorithm, based on direct density ratio, is used to detect a step change in a covariate. A change in covariate may be driven by a fault event, for example ingestion of foreign object or debris.

Figure 6.9 illustrates how detecting the change in a covariate level correlates with the rapid change in TGT margin. The top subfigure illustrates the real TGT margin and its "expected" ground truth, represented by cross, "+", and solid line, respectively. For every engine, to estimate "expected" ground truth, linear regression has been used on the whole dataset, based on consultation with industrial peers. The vertical dash line represents where the degradation change occurs. The first stage, between flight cycle 0 and 45, is a normal deterioration, whilst the second stage, after approximately flight cycle at 45, is a more rapid deterioration. The change in the covariate is shown by the middle subfigure, whereas the score of CPD is illustrated by the bottom subfigure.

In order to obtain the change information for each degradation data, CPD algorithm is applied. The main prognostic algorithm, BR-3-VB, uses two types of prior parameters for this problem. The first prior parameters are calculated based on two stages of all available degradation data, whilst the second prior parameters are computed based on all rapid deterioration only. The CPD algorithm functions to detect the anomaly/significant change in the covariate level. As shown in figure 6.7, if CPD algorithm detects the change in the covariate level, i.e. the score of CPD algorithm crosses the defined threshold, the prognostic model can be reconfigured based on second type of prior parameters.

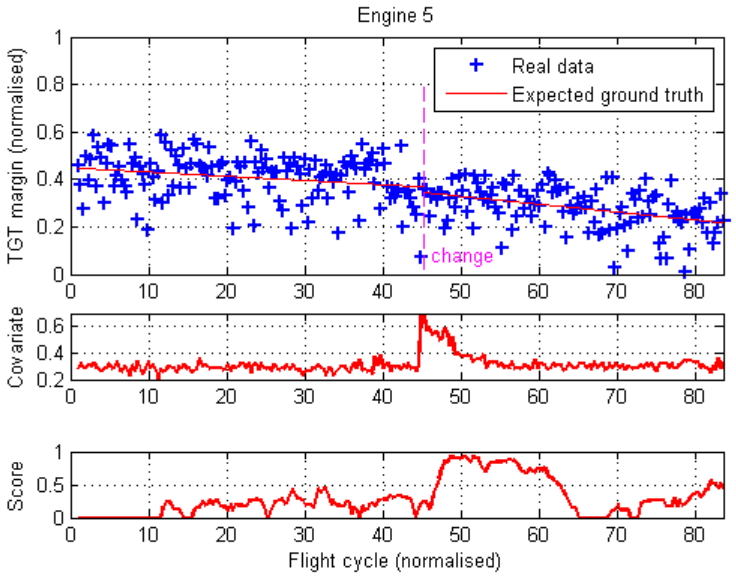
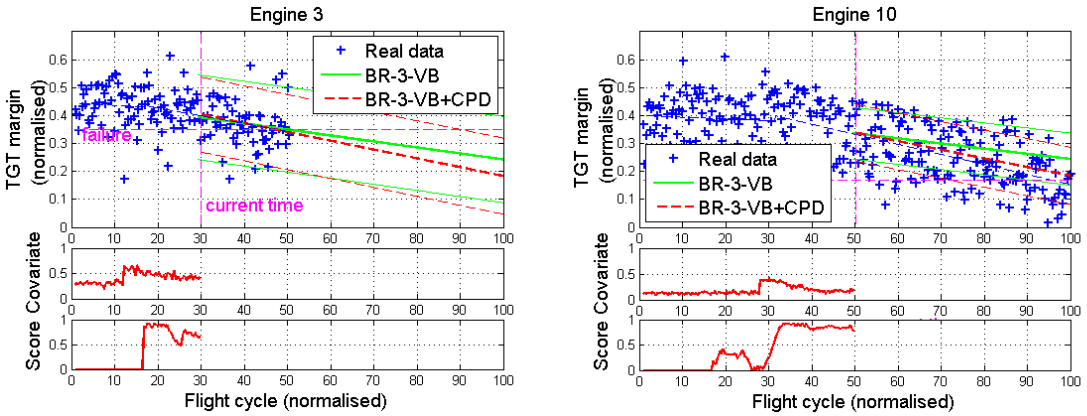


Figure 6.9: Change point detection shows the correlation between the change in covariate and TGT margin.

A score threshold is determined based on off-line implementation of CPD into the covariate data. From several experiments, this threshold can be specified. Furthermore, other CPD parameters, such as n, k and α, can be selected heuristically based on some simple experiments on the available data. In this case study, the score threshold is set to be 0.6, whereas the selected parameters are $n = 10$ and $k = 15$. The choice of these parameters is suitable to create "moderate" Hankel matrix. This choice is not too small, so that it may capture change characteristics in the data and it is also not too large with the result that the available data would be sufficient enough to form Hankel matrix. The score will be bound at $1/\alpha$, so that we select this to be $\alpha = 0.5$.

This case study attempts to compare the prognostic algorithm performances, with and without the involvement of CPD algorithm. Ten extreme scenarios of real TGT margin degradation are selected, where all degradation signals contain various degrees of rapid change in degradation slope. In this scenario, different failure thresholds are selected to ensure the "expected " ground truths of degradation cross the thresholds, because most real data are right censored.

Figure 6.10 shows trajectory predictions of two different slope changes. The cross,



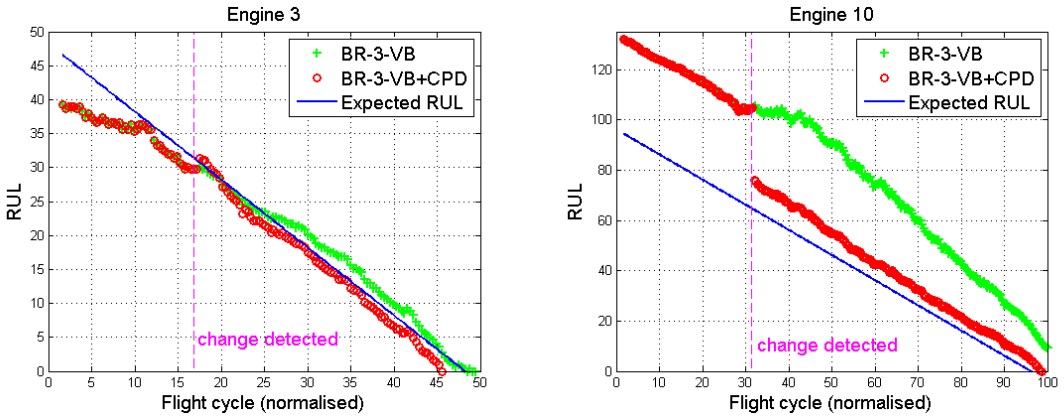
(a) Engine 3

(b) Engine 10

Figure 6.10: Trajectory predictions of BR-3-VB and BR-3-VB+CPD. The left figure is a degradation with slow decay and the right figure is degradation with rapid decay.

"+", is the real TGT margin data. The thick straight solid and dashed lines are BR-3-VB and BR-3-VB with CPD algorithm (BR-3-VB+CPD), respectively. The vertical and horizontal dashed lines are current time and failure threshold, respectively. The figure 6.10a shows when there is slow decay in degradation, both BR-3-VB and BR-3-VB+CPD are still able to predict well. The reason is because BR-3-VB alone has a common prior distribution which has information across the engine's population, described in chapters 4 and 5. However, when the rapid decay occurs as shown in the figure 6.10b, BR-3-VB alone cannot predict well the degradation, whereas the BR-3-VB+CPD is able to maintain the prediction close to the "expected" ground truth. If there is a step change in covariate, CPD algorithm provides information to BR-3-VB for reconfiguring the prognostic prior parameters to enhance RUL estimation.

Figure 6.11 demonstrates two RUL metrics for these scenarios. The solid line, cross, "+", and the circle, "○", represent "expected" RUL, BR-3-VB and BR-3-VB+CPD, respectively. The figure 6.11a shows RUL metric for a slow decay problem. This figure emphasises that for the slow decay problem, involving CPD might not be crucial because both predictions provide similar performance. However, as illustrated in the figure 6.11b, for a rapid decay problem, BR-3-VB alone converges very slowly to the "expected" RUL, whilst BR-3-VB+CPD can be reconfigured when there is a change in covariate, resulting in better RUL estimation.



(a) Engine 3

(b) Engine 10

Figure 6.11: RUL metrics for BR-3 and BR-3 with CPD for two different degradation scenarios. The left figure is RUL metric for slow decay whereas the right figure is RUL metric for rapid decay.

In order to prove the consistency of the proposed concept, residuals of "expected" RULs for 10 different engines are computed. These residuals can be represented as two box plots, illustrated by figure 6.12. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers. The top subfigure shows BR-3-VB whereas the lower subfigure shows BR-3-VB+CPD. At beginning of the prediction, between approximately flight cycles 0 until 30, both box plots demonstrate similar performance. However, after around flight cycles 30, BR-3-VB and BR-3-VB+CPD start to vary due to the change in degradation slope. It can be observed that, after approximately flight cycle 30, the BR-3-VB alone has larger variations than BR-3-VB+CPD, because BR-3-VB does not handle well rapid decay problem, which occur in some degradation data. This result proves that the proposed concept to incorporate CPD algorithm into the prognostic approach is beneficial to improve RUL estimation.

6.3.2.2 Handling maintenance event in a recoverable system

As described in section 1.2.3.6, maintenance action, such as compressor washing, will recover degradation performance. The main prognostic algorithm requires information about when this event occurs, in order to reset the algorithm. However, as described

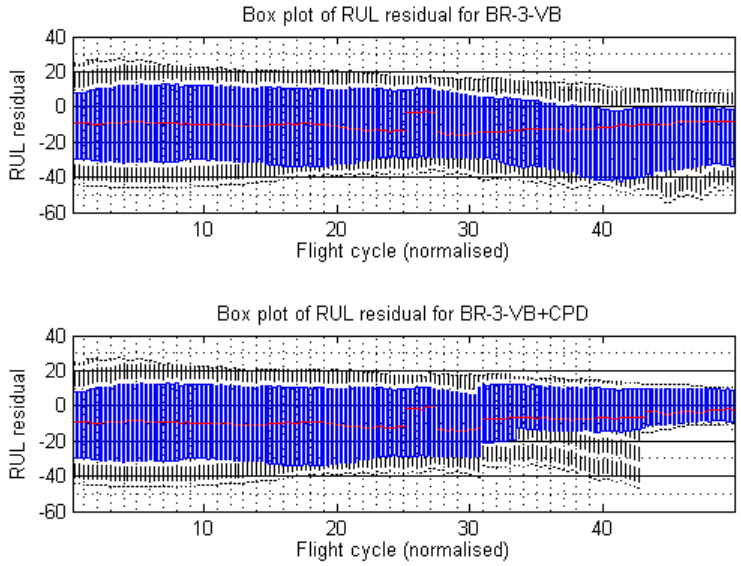


Figure 6.12: Box plots representing comparison of overall performance between BR-3-VB and BR-3-VB+CPD.

earlier, the maintenance events are not always known in practice. Therefore, CPD is applied directly to TGT margin for detecting abnormality in the data, as shown in figure 6.7.

As explained in section 6.3.2.1, some CPD parameters, such as n, k and α, can be tuned heuristically based on some simple experiments on the available data and threshold can be chosen in consultation with the subject experts. In this case study, the selected parameters are $n = 10$, $k = 15$, $\alpha = 0.5$ and threshold value is chosen to be 0.7.

Figure 6.13 demonstrates an example of TGT margin and the score of CPD. The top subfigure is TGT margin data and its "expected" ground truth, represented by cross, "+", and solid line, respectively. It can be observed that the TGT margin increases on flight cycle 40, indicates that maintenance has been performed. The lower subfigure shows a change point score based on PE_α(symmetric). An increasing score, > 0.7, can be observed at approximately flight cycle 45, indicating the detection for the maintenance event. The delay occurs because the used CPD algorithm, RuLSIF, is based on retrospective algorithm.

Figure 6.14 shows a RUL metric for a recoverable TGT margin data. The solid line

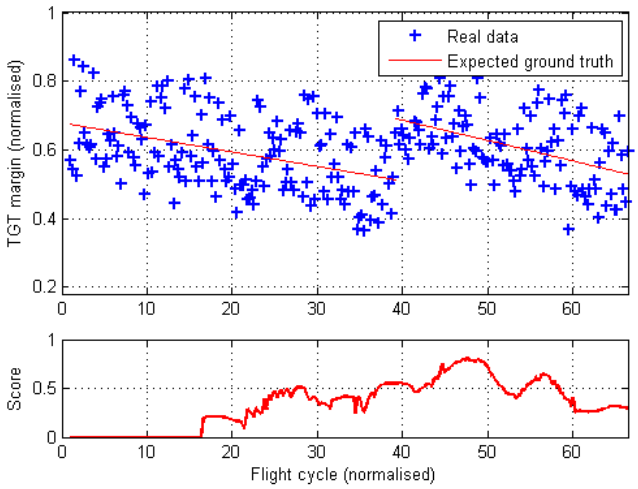


Figure 6.13: CPD is able to detect significant changes due to unknown maintenance event took place.

is the "expected" RUL and the circle, "○", is the estimated RUL, based on BR-3-VB. A significant increase in the "expected" RUL is observed around flight cycle 40 which indicates the occurrence of a maintenance action. CPD algorithm detects the change approximately at flight cycle 44. Once the maintenance action has been detected, the prognostic algorithm is then reset to restart the new prediction. The small delay in detection should not be a big problem, because gas turbine engine prognostics is an off-line process, where there is an allowed tolerance time for decision making. Guaranteed accuracy and robust detection are more important for the success of prognostic programme.

6.4 Summary

This chapter presents how prognostic performance can be improved by utilising the information about irregular events, such as maintenance event, slope change in degradation as well as including information from different levels of gas turbine engine into the prognostic framework.

A generic methodology of integrated prognostic algorithm is introduced. This allows prognostic algorithm to be reset due to maintenance action, and its prior param-

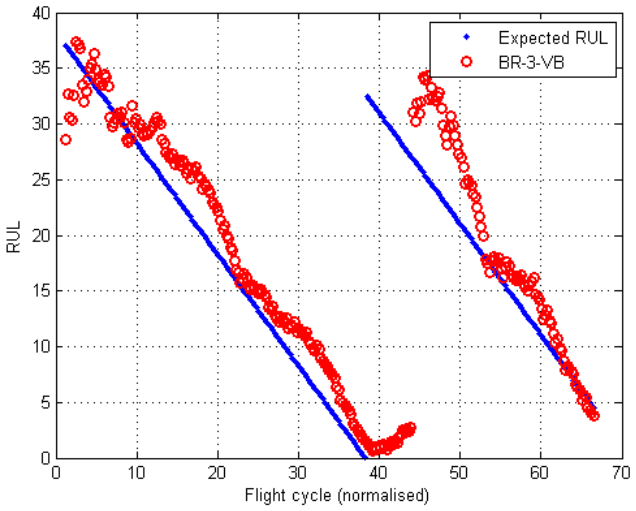


Figure 6.14: Illustration RUL metric where CPD algorithm detects the change around flight cycle 44 to reset the prognostic algorithm.

eters can be reconfigured when there is a change in degradation slope. The proposed methodology proves its capability through the use of synthetic data.

However in practice, detecting any slope change in degradation data is difficult because the data are typically very noisy. In reality, maintenance events are also often unknown because airlines, which carry out maintenance, do not always provide such information to organisation, which perform fleet management. The uncertainty surrounding maintenance action would impact in determining stopping and starting in prognostic analysis.

To overcome these issues, a sophisticated integrated prognostic approach is introduced, combining Bayesian methods with change point detection (CPD) algorithm. BHM with variational Bayes, termed BR-3-VB, is used as the main prognostic algorithm due to its capability in dealing with uncertainty and accommodating optimally multiple fleets of engine. CPD algorithm, based on estimating directly the relative density ratio, plays an important role to detect any changes in degradation directly as well as its covariates from different levels of hierarchy in gas turbine engine architecture. The strength of the CPD concept lies in the fact that it estimates the relative density ratio between two segments of the time series data directly rather than estimating the probability density function separately, which is a hard task, and then comparing it for

divergences.

In the first real case study, CPD algorithm is used to detect the change in a covariate, which affect degradation's slope and this information is then utilised to accordingly reconfigure the prognostic model for enhancing the quality of RUL estimation. Furthermore, in the second real case study, CPD algorithm is applied directly to TGT margin data to discover when the maintenance action took place. This information supports the main prognostic algorithm to restart the prediction. The results for both case studies reveal the significant improvement in quality of RUL estimation.

In conclusion, the proposed integrated prognostic concept proves to be promising to be applied in a complex hierarchical system. This method is able to detect any changes or faults in multiple covariates (e.g. vibration, ambient temperature) at any level of the system's hierarchy, including sub-system as well as component level. The challenge is to select optimally the information from multiple covariates which may affect the degradation of health parameter indirectly as well as performance of prognostics algorithm.

The future direction of this method would be to design an integrated prognostic algorithm which can select intelligently the most dominant covariates and fuse the information from those promising covariates together, to inform intelligently the main prognostic algorithm, BR-3, as well as optimally using the multiple fleets of engine for enhancing RUL estimation of gas turbine engine.

Chapter 7

Conclusion and Future Work

This chapter summarises several major contributions in this thesis and provides several suggestions for future works.

7.1 Main Contributions

The main objective of this research is to develop generic data-driven prognostic algorithms with applicability in various scientific and engineering domains, where in this particular work, the developed techniques are applied to the degradation data obtained from gas turbine engines. The developed algorithms aim to deal with various challenges, arise in complex systems, including gas turbine engines. Selecting an appropriate algorithm for a particular application is crucial to the ultimate success of a prognostic programme. Therefore, this requires a good understanding of challenges associated with a specific application.

A gas turbine engine can be considered as a complex system because it comprises multiple interacting subsystems. Engine degradation which takes place at the system level (e.g. TGT margin), may also be influenced by the changes/faults that occur at subsystem or component level. Therefore, the first challenge in the gas turbine engine prognostics is an adversity to construct a physics-based approach that mimics the dynamics of the systems. Knowledge-based approaches also become harder when the experts need to specify rules for a complex system with numerous of interrelating failure modes. On the other hand, a considerable volume of health signal data is available from modern engines because there are large number of engines to be monitored with

multiple EHM sensors, which are attached on each engine. Hence, this research focuses on data-driven based prognostic approaches.

Further complication that arises is the inexistence of ground truth, i.e. assets are never allowed to fail in service, so that RUL can only be estimated. Therefore, a main health index, TGT margin degradation data, is difficult to use, directly, for testing and validating prognostic algorithms. In this thesis, we introduce a methodology to use synthetic data. The synthetic data of known properties, which mimics noisy degradation data, are generated. This methodology is very effective to validate the performance of the developed algorithms, because in the synthetic data, the model is known and its parameters can be controlled. Subsequently, these algorithms are applied to the real degradation data, for validation.

In a gas turbine engine, TGT margin represents the overall health of the system. This health parameter is usually used to monitor the gas path degradation of the engine to detect the changes in performance for each engine and to indicate the need for inspection/maintenance. Thus, TGT margin is forecast for RUL estimation of the engine. However, there is a large uncertainty associated with TGT margin data as it gets corrupted with noise due to various reasons, including gas turbine design, manufacturing, ambient and environmental condition, operating condition, duty mission, maintenance action, etc. Large uncertainty in the data may cause inconsistency in prognostic prediction, especially when there is little data available. This research devotes on developing advanced Bayesian approaches to deal with large uncertainty issue, where they should be able to enables variation and uncertainty to be quantified, mainly by using distribution instead of fixed values in risk assessment.

Bayesian method is transparent in their assumptions, where it provides a clear view of the rules that govern the relationships that make predictions possible. For the validation purpose, the transparent solution is important in the field of safety-related applications. Bayesian method also allows incorporation of previous knowledge/experience in a coherent way and avoids over-fitting problems. Furthermore, Bayesian techniques have been utilised widely and successfully in multidisciplinary fields related to prognostics, such as reliability engineering, survival analysis and forecasting.

In this thesis, there are four main contributions related to Bayesian approaches to cope with further challenges in gas turbine engine prognostics. Figure 7.1 illustrates the main contributions in this thesis, discussed in four chapters. Chapters 3-5 discuss

mainly about various Bayesian prognostic techniques and chapter 6 describes combination between Bayesian prognostics with CPD algorithm (event detection). The detailed contributions will be outlined in the following subsections.

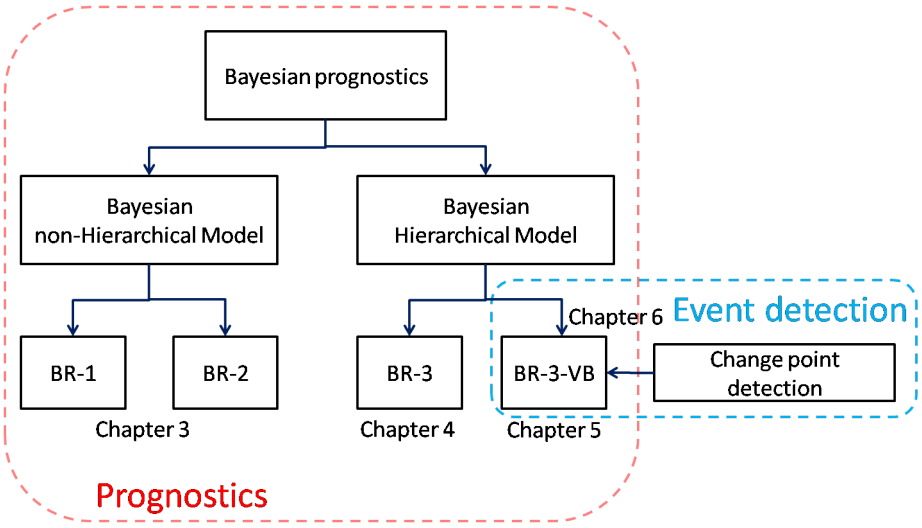


Figure 7.1: Illustration of main contributions in the thesis. Chapters 3-5 discuss Bayesian prognostics whereas chapter 6 describes combination between Bayesian prognostics with CPD algorithm (event detection).

7.1.1 Bayesian Regression 2

This research had been initiated by adopting main concept of a well established Bayesian prognostic method, developed by Gebraeel [2006]; Gebraeel et al. [2005], which can be called Bayesian Regression 1 (BR-1). This Bayesian method has an analytical solution because it involves conjugate Gaussian prior distribution. The advantages of this method are that the parameters in the Gaussian prior are relatively easy to determine based on available data and the posterior solution can be computed deterministically.

However, BR-1 assumes that the variation in observable health index is constant over the life of asset, i.e. the estimation errors are assumed to have a fixed noise variance. This assumption may lead to inflexible uncertainty in capturing noise characteristics. To overcome this limitation, we propose an alternative Bayesian prognostic method, named Bayesian Regression 2 (BR-2). Both approaches produce closed-form solution, desirable for real-time, deterministic computation. The main contribution in BR-2, lies

in modelling of the observation noise as a random variable allowing for the variation in noise level for each iteration.

To investigate the strengths of the proposed approach, both methods are tested on synthetic degradation data, where they present predictive results within well defined uncertainty bounds. This demonstration reveals that BR-2 is promising to be implemented when degradation signals have different variation between units and they also have high level of noise. This scenario is likely to occur in reality, such as TGT margin degradation.

For validation, BR-1 and BR-2 are applied to real TGT margin data prognostics in the second case study. The outcome proves that the synthetic data conclusion is preserved in the real case study. BR-2 demonstrates the improvement of the results in terms of accuracy and consistency for the case when degradation signals have different variations between units with high level of noise due to the noise inclusion in Bayesian model.

7.1.2 Bayesian Regression 3 using Gibbs sampler

A considerable number of health signals as a fleet of assets is available from modern civil aerospace gas turbine engines. It is a challenge for prognostic algorithms to use, optimally, data available from multiple fleets of engine for estimating the RUL of a specific engine. The state-of-the-art prognostic algorithms are able to update degradation models with data collected from individual assets with in-service measurements. However, the ability to accommodate the heterogeneous fleet into a single prognostic model is not well studied. BR-1 and BR-2 can be considered as Bayesian non-Hierarchical Model (BnHM), where they do not utilise, optimally, data available from multiple engines.

To overcome the above issue, Bayesian Hierarchical Model (BHM) prognostics, named BR-3 (with Gibbs sampler), is introduced in this thesis. BR-3 has its own hierarchy, where the model coefficients vary at more than one level. This capability allows the models to obtain information from both prior distributions: an individual prior, related to a particular unit and a common prior, related to the whole fleet of units. This type of model has been successfully used in many fields, including biostatistics, reliability engineering, econometrics, machine learning, etc. However, the author has not found any

literature addressing about the use of the hierarchical model in prognostic applications.

BR-3 is appropriate to maximise the use of the assets' data by accommodating them into a hierarchical model. In this way, the health information can be shared between the engines in order to enhance RUL estimation of the engines.

Nevertheless, according to several literatures in Bayesian statistics, it is difficult to specify prior parameters for BHM. A well known method is to use a non-informative prior for IG prior, called a Jeffreys prior. This choice has been widely used in many applications. However, recently Gelman [2009] stated that this non-informative prior can sometimes have strong and undesirable implications. To overcome this issue, a methodology to produce "moderately" informative prior distribution is introduced in this thesis. This methodology and a well-known Jeffreys prior choice are compared using synthetic data, where the results show that the proposed methodology outperforms this non-informative prior. Another analysis is also conducted to select a crucial scaling factor in a prior parameter to ensure BR-3 can produce appropriate prognostic performance.

In order to demonstrate the strengths of BR-3, a comparative study between BR-2 and BR-3 is carried out using synthetic data as well as real degradation data for evaluating their relative performance. In general, there is no substantial difference between BR-2 and BR-3, applied to standard degradation data, such as used in chapter 3. However, BR-3 outperforms BR-2 in some extreme cases in degradation data, such as slope change and very high noise level in degradation signal. BR-3 is able to share the mutual health information owing to common prior (level 2) and this is then combined with individual prior (level 1). It can be concluded that BR-3 is promising to be implemented in complex system prognostics, where there may be some extreme scenarios, occurring in degradation signals.

7.1.3 Bayesian Regression 3 using variational inference

BR-3 is based on a complex Bayesian model, where it does not have a closed-form solution, which requires the use of the Gibbs sampler for computing conditional posterior distributions. However, Gibbs sampler may not be attractive for use in prognostic field owing to high computational cost and the level of expertise needed for determining some crucial parameters (diagnostic convergence). This type of sampling should also

be restarted in each prognostic iteration, which would lead to computational inefficient. A sequential type of MCMC may solve this problem, but would require an even higher level of expertise to run it.

In practice, these issues result in less autonomy for the prognostic system and may require big investment in powerful computational resources, e.g. supercomputer facility, which ultimately it might not generate prognostic benefits in term of return of investment. To cope with these crucial issues, an approximate inference based on variational Bayes for BR-3, called BR-3-VB, is derived and developed, which delivers an approximate but deterministic solution. The derivation includes the solutions for posterior and predictive distributions as well as variational lower bound.

A comparative study is conducted, where BR-3 and BR-3-VB are firstly tested on synthetic data, used in chapter 4. The results demonstrate that BR-3 and BR-3-VB have similar performance in RUL estimation. In the second case study, both approaches are applied to real TGT margin data for validation. The outcome proves that the synthetic data conclusion is preserved in the real case study. The quality of approximation is shown to be satisfactory with respect to prediction performance, computational speed and ease of use. Hence, variational Bayes is a promising inference to be an alternative to Gibbs sampling for an analytically intractable Bayesian model. In particular, BR-3-VB has a great potential to be implemented in complex system prognostics as it produces similar capability with BR-3 in terms of accuracy as well as consistency and also delivers deterministic solutions, suitable for autonomous prognostic system.

7.1.4 An integrated prognostic approach

In gas turbine engine degradation, most of the degradation patterns are nearly linear, however there are cases where the rates of degradation may be non-linear. In the second case, the engine performance may degrade at approximately constant rate lasts for a period of time, followed by an increase in rate. This occurs due to various factors, including a step change in covariates as well as fault modes. Another issue takes place when engine performance recovers due to maintenance action. For accurate prognostics, knowledge of maintenance actions which affect the rate and state of degradation is crucial.

A methodology is proposed where prognostic algorithm can be reset when mainte-

nance is performed and the prior parameters are reconfigured when there is a change in degradation slope. This concept is proved to be efficient through the use of synthetic data.

However, in reality, it is difficult to detect directly any slope change in degradation because they are typically very noisy. Maintenance events are also often difficult to obtain and incorporate. The maintenance actions are performed at geographically diverse locations by organisations independent to those performing fleet management, these factors lead to uncertainty in the maintenance state of the asset. The uncertainty surrounding maintenance actions impacts the ability to accurately predict the degradation.

In order to overcome the aforementioned issues, a sophisticated integrated prognostic approach is introduced, combining a Bayesian data modelling technique with information theoretic change point detection (CPD) algorithm. BR-3-VB is used as the main prognostic algorithm because it is deterministic and is capable in dealing with uncertainty and accommodating optimally multiple fleets of engine, whilst CPD algorithm, based on estimating directly the relative density ratio, plays an important role to detect any changes in degradation directly as well as its covariates from different levels of hierarchy in gas turbine engine architecture. The strength of the CPD concept lies in the fact that it estimates the relative density ratio between two segments of the time series data directly rather than estimating the probability density function separately, and then comparing it for divergences.

In the first real case study, CPD algorithm is used to detect the change in a covariate, which affect degradation's slope and this information is then utilised to accordingly reconfigure the prognostic model for enhancing the quality of RUL estimation. Furthermore, in the second case study, CPD is applied directly to TGT margin data to discover when the maintenance action took place. This information supports the main prognostic algorithm to restart the prediction. The results for both case studies reveal the significant improvement in quality of RUL estimation.

The proposed integrated prognostic concept proves to be promising to be applied in a complex hierarchical system. This method is able to detect any changes or faults in multiple covariates (e.g. vibration, ambient temperature) at any level of the system to then inform intelligently main prognostic algorithm to update the belief about degradation.

7.2 Suggestions for Future Work

Despite the proposed prognostic algorithms prove to be promising, there are still a number of research areas, requires to be improved. This section provides several suggestion for future works.

7.2.1 Health index

As described in section 2.2, a successful prognostic programme does not solely rely on a prognostic algorithm. In overall prognostic process, many other factors also play a significant role in enhancing prognostic performance. These factors include sensor system, data acquisition, feature extraction and diagnostics.

In gas turbine engine's case, the take-off TGT margin data is a feature of health index, representing overall health of a gas turbine engine. However, the supplied TGT margin data are very noisy and do not always show consistency[1]. As a result, the developed Bayesian approaches will produce large uncertainty in failure-time distribution, leading to difficulty in decision making for maintenance.

It is crucial to improve the quality of TGT margin extraction. This improvement should produce more "reasonable" degradation signal, where degradation trend should be consistent and variation in data should be shrunk. Ultimately, this would contribute better prognostic performance. This task should involve multidisciplinary research, including sensor systems [MacIsaac and Langton, 2011] and feature extraction [General-Electric, 2008]. Since these issues are not author's expertise, these will not be discussed in detail how this research can be carried out further.

7.2.2 Ground truth

TGT margin is a main health feature to be forecast, considered degradation signal (section 1.2.2). In our work, the degradation signal is extrapolated, in the form of predictive distribution, until it crosses action warning threshold first and it is then converted into a failure-time distribution for estimation RUL.

[1]Proper degradation data should show consistent degradation behaviour (monotonic decrease/increase) [Saxena et al., 2008a].

On the other hand, there are several prognostic techniques which utilise failure-time data directly into a prognostic algorithm, such as Caesarendra et al. [2010] and Widodo and Yang [2011a] for bearing applications. They assumed directly this data type as a likelihood function. This strategy leads to a full Bayesian analysis for failure-time distribution which might have an advantage over degradation data.

However, the main obstacle in TGT margin data is the absence of ground truth. The ground truth data is very crucial to produce true failure-time data as well as an appropriate algorithm validation. In this work, based on consultation with industrial peers and expert matters, linear regression method is applied to TGT margin data and it assumes the mean of the regression model as the "expected ground truth". Nevertheless, this practice does not have a strong scientific foundation. Specific research in determining ground truth is strongly required, where the ground truth is rarely available in many situations, especially in a complex system [Saxena et al., 2008a].

7.2.3 Covariates

As described in section 1.2.3.1, multiple patterns may exist in gas turbine engine degradation, due to various factors, including covariate effects. In the proposed prognostic model, such as described in equation (5.7), a design matrix component, Φ, can be used to allow naturally covariate inclusion, such as described in Tobias [2001]. The inclusion of covariates is expected to improve the prediction.

In statistics, there are two types of covariate: time-independent covariate and time-dependent covariate [Davidian and Giltinan, 1995]. A time-independent covariate is a covariate where it remains fixed whereas a time-dependent covariate reflects the phenomenon that a covariate varies over time through whole observation/experiment. In the case of gas turbine engine prognostics, examples of time-independent covariate are fault, engine type and operator (airline), whist examples of time-dependent covariates are ambient temperature and engine pressure ratio.

In our case, flight cycle is used as an input of prognostic model. The flight cycle can be classified as a time-dependent covariate because it varies over time, which affects degradation in a gas turbine engine. In addition, "total hours flown" may be treated as another time-dependent covariate. For instance, an engine, used from Jakarta to London, may have different degradation with, another engine, used from Amsterdam to London.

Considering "total hours flown" as a covariate, may improve RUL estimation.

Basically, including time-dependent covariates is a difficult task, because when a main health index is extrapolated, it is also necessary to forecast time-dependent covariates. Instead of constructing another forecasting method for the covariates, the most effective method is to use an existing forecasting model to support a prognostic model, such as weather forecasting, flight planning, etc. For instance, ambient temperature must be forecast together with TGT margin for prognostics. Here, the information about location and annual temperature in the airports, where an aircraft takes off, can be used to substitute extrapolation of ambient temperature.

On the other hand, time-independent covariates are relatively easy to be included in the prognostic algorithms because forecasting of covariates is not required. The covariates can be included directly into the component Φ in the prognostic model. For example, an operator (airline) is a time-independent covariate, where airlines A, B and C can be distinguished in prognostic model by tagging them to be numbers 1,2 and 3, respectively. This may assist prognostic model to differentiate airline and estimate the RUL based on their degradation behaviour.

Furthermore, as described in chapter 6, a step change/fault can be detected through CPD on covariate level. This scenario is also classified as a time-independent covariate's problem. The detection is then used to reconfigurable the main prognostic model to enhance RUL estimation.

There are numerous covariates in a gas turbine engine, including ambient temperature, pressure, humidity, flight length, etc [Müller et al., 2010]. Other diagnostic/detection techniques should be considered to be tested at different level, subsystems as well as components, to identify more changes/faults. In future, the research direction should be to design an integrated prognostics algorithm which can select intelligently the most dominant covariates and fuse the information from those promising covariates together, to inform intelligently the main prognostic algorithm, BR-3-VB, as well as optimally using the multiple fleets of engine for enhancing RUL estimation of gas turbine engine. Further research is required to be conducted to investigate the benefits.

7.2.4 Flexible model

Another way to capture multiple degradation patterns is by applying a flexible model. Prognostics is a time-series problem, where time-series can be defined as a sequence of data assigned to specific moment in time. As described in section 2.3.3, prognostics has a very strong relationship with forecasting. Here, Bayesian forecasting methods can be adopted as a prognostic algorithm.

BR-3 and BR-3-VB are based on a hierarchical modelling, which can be extended to be a dynamic model, known as panel Bayesian vector autoregressive (BVAR) model [Canova and Ciccarelli, 2013; Giannone et al., 2012]. A hierarchical model has also been advocated by the first proponents of BVARs [Canova, 2007; Del Negro and Schorfheide, 2011; Doan et al., 1984; Giannone et al., 2012; Sims and Zha, 1998].

In economic sciences, Christopher A. Sims[1] was one of the main promoters of the use of BVAR in empirical macroeconomics for evaluating economic policies. In this field, the prediction is normally made for a short term. For example, Gross Domestic Product (GDP) is forecast every one year, where data point is observed every year for a long period (many years) [Tobias, 2001]. On the other hand, prognostics requires a long term prediction. For instance, the data point (TGT margin) is observed every day in gas turbine engine prognostics, but the prediction horizon needs to be forecast until it crosses the threshold, where the extrapolation may be within months or even years in the future.

Applying the dynamic model in the form of BR-3 might provide the benefits in capturing non-linear degradation behaviour. However, these models are effective for short term predictions, such as forecasting problem, but less reliable when it is used for long-term predictions, such as prognostics, due to dynamic noise, their sensitivity to initial system conditions and an accumulation of systematic errors in the predictor [Sikorska et al., 2011].

Another suggestion is to apply a complex model, such as Gaussian Process (GP). GP is a probabilistic approach for learning in kernel machines, which is mathematically equivalent to many well known models, including large neural networks (under suitable conditions), spline models and Bayesian linear models [Rasmussen and Williams, 2006]. In prognostics, GP model has been used to estimate RUL in battery and crack

[1] Christopher A. Sims is a Nobel memorial prize winner in economic sciences in 2011.

length of metallic aircraft structural components by Goebel et al. [2008b] and Mohanty et al. [2009], respectively. However, their proposed GP models are "only" equivalent to BnHM, which do not have capability to share information across units.

In order to express BHM (e.g. BR-3) as a GP model, a hierarchical GP model should be used, such as described in Menzefricke [2000]. The main issue with the hierarchical GP model is the need of sampling methods to make an inference, such as a hybrid Monte Carlo, which have drawbacks (in prognostic practice) as described in section 5.1. The main challenge is to make an approximate inference to this model. Alternatively, our algorithm, BR-3-VB, should also be able to capture a complex degradation behaviour, by applying different choice of basis functions in the design matrix component, $\mathbf{\Phi}$.

7.2.5 Hybrid approach

The trends of all characteristic parameters are diversified and difficult to be predicted by using a single prediction method in the real-world prognostic processes [Peng et al., 2010]. Thus, a hybrid approach should be adopted to combine the strengths of individual prognostic approaches in order to improve RUL estimation [Pecht and Jaai, 2010; Peng et al., 2010]. This can be produced by combining either physics-based, knowledge-based or data-driven approaches.

For example, Brotherton et al. [2000] combined a data-driven with knowledge-based approaches, to dynamically linked ellipsoidal basis function neural network with rule extractors, respectively for gas turbine engine prognostics.

Another example was proposed by Liu et al. [2012]. They integrated the strengths of the data-driven prognostic method and the model-based particle filtering approach in predicting the RUL of lithium ion batteries through electrochemical impedance spectroscopy tests. In the proposed methodology, particle filtering is applied for system state estimation in parallel with parameter identification of the prediction model (with unknown parameters) based on Bayesian learning. Simultaneously, a data-driven predictor is employed to learn the system degradation pattern from history data so as to predict system evolution (or future measurements).

In our case, there is a possibility to combine knowledge-based approach with our developed data-driven approaches. For instance, expert knowledge (e.g. experienced engineers and technicians) is highly valuable to be considered [O'Hagan and Oakley,

2004]. An expert elicitation should be developed and this can be combined with the predictive or failure-time distributions to provide better RUL. Figure 7.2 illustrates the implementation of combining prognostic using Bayesian approach and elicitation experts' judgements. This combination should increase the belief of failure-time distribution, by shrinking its distribution.

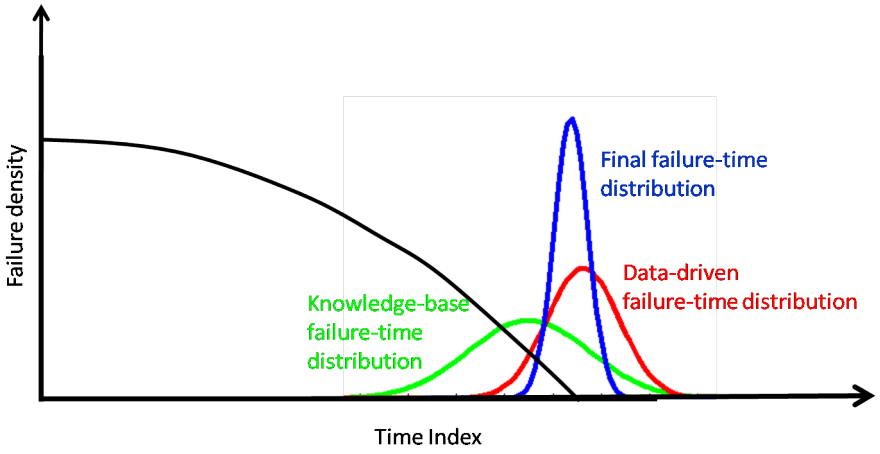


Figure 7.2: Illustration of the implementation of combining prognostic using Bayesian approach and elicitation of expert knowledge.

Similar methodology may also be applied for a brand new engine, when there is no in-service data available, lead to the difficulty in determining prior parameters. A new engine is normally examined under reliability testing in the laboratory prior to be launched into market. The reliability data can be utilised to specify prior parameters. In addition, knowledge-based approach can be used here to determine prior parameters by eliciting from experts' judgements. This methodology can also be classified as a hybrid approach, where data-driven and knowledge-based approaches are combined.

As above explanation, there would be a significant benefits, fusing together all prognostic approaches. The future of prognostic algorithm would lie on a hybrid approach to obtain the strengths of each approach and provide robust RUL estimation.

7.2.6 Post-prognostics

7.2.6.1 Reconfigurable control system

The ultimate goal of prognostics research should be to make links between prognostics and control systems to provide extra benefits. Some studies have been carried out by National Aeronautics and Space Administration (NASA) to investigate the benefits including prognostics in control loop [Brown et al., 2009; Garg, 2004; Tang et al., 2008]. This inclusion allows not only fault accommodation, but also fault mitigation via proper control actions based on short term prognostics, and moreover, the establishment of a long term operational plan that optimises the utility of the entire system based on long term prognostics. For example, the consideration of prognostics at component level, control reconfiguration provides a unique opportunity to manage component life via control actions [Tang et al., 2008].

7.2.6.2 Cost-benefit analysis

As described in chapter 1, prognostics delivers many benefits, but prognostic implementation requires consideration and planning for integration into new and existing systems, operations, and processes [Feldman and Sandborn, 2008]. Prognostics costs money in terms of acquisition and installation costs, implementation costs, and changes in business practices [Sun et al., 2012].

The cost of re-design of host product can be an expensive investment. For example, the original cables need to be re-wired to supply power to the sensor for deploying a sensor and microprocessor on a ball bearing or gearbox. The casting must also be re-designed to take in the sensor and protect it from the environment. These implementation costs need to be accounted for. Another example is described in chapter 5, where a big investment needs to be spent in powerful computational resources to run prognostic algorithms based on sampling method. This investment should be considered as well in term of prognostic cost benefits. The adoption of prognostics must provide a significant advantage in order to add value for the maintenance process. If there is no economic benefit, system vendors may not wish to implement prognostics [Sun et al., 2012]. Cost-benefit analysis and quantitative assessment are therefore essential for assessing the effectiveness of prognostics [Banks et al., 2005].

There are several financial metrics that can be used in a cost benefit quantitative analysis, including net cash flow, cumulative cash flow, payback, return on investment, net present value, and internal rate of return [Banks and Merenich, 2007]. Among all these metrics, return on investment is one of the most selective metrics [Sun et al., 2012]. In general, return on investment is the ratio of gain to investment [Feldman et al., 2009], which tells us the rate of return on the investment in prognostics, which enables the investment in prognostics to be compared with other competing investments [Saxena et al., 2008a; Wang and Pecht, 2011].

This metric involves an analysis of the cost avoidances made possible by using prognostics technology against the costs associated with the development, manufacturing, installation, and implementation of prognostic technology in selected systems [Sun et al., 2012]. Therefore, it is important to carry out more studies on the life cycle return on investment to the implementation of prognostic technologies.

Appendix A

Probability Distributions

This appendix describes briefly the main properties of some probability distributions, used in this thesis. The used notations follow the same notations in (Bishop, 2006).

A.1 Gamma distribution

The gamma distribution is a two-parameter family of continuous probability distributions, given by:

$$\text{Gam}(\tau|a,b) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} e^{-b\tau} \tag{A.1}$$

$$\mathbb{E}[\tau] = \frac{a}{b} \tag{A.2}$$

$$\text{var}[\tau] = \frac{a}{b^2} \tag{A.3}$$

$$\mathbb{E}[\ln \tau] = \psi(a) - \ln b \tag{A.4}$$

$$H[\tau] = \ln \Gamma(a) - (a-1)\psi(a) - \ln b + a \tag{A.5}$$

where $\mathbb{E}[\tau]$, $\text{var}[\tau]$ and $H[\tau]$ are mean, variance and entropy of Gamma distribution, respectively. Parameters $a, b > 0$, $\tau \geq 0$ and $\psi(.)$ is digamma function (Abramowitz

and Stegun, 1965), given by:

$$\psi(a) \equiv \frac{d}{da} \ln \Gamma(a) \tag{A.6}$$

A.2 Gaussian distribution

The Gaussian (Normal) distribution is a continuous probability distribution, which is also known as the normal distribution. It is defined by the formula:

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{D/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu) \right\} \tag{A.7}$$

$$\mathbb{E}[\mathbf{x}] = \mu \tag{A.8}$$

$$\text{cov}[\mathbf{x}] = \Sigma \tag{A.9}$$

$$H[\mathbf{x}] = \frac{1}{2} \ln\, |\Sigma| + \frac{D}{2}(1 + \ln(2\pi)) \tag{A.10}$$

where $\mathbb{E}[\mathbf{x}]$, $\text{cov}[\mathbf{x}]$ and $H[\mathbf{x}]$ are mean, covariance and entropy of Gaussian distribution, respectively.

If we have a marginal Gaussian distribution for $\mathbf{x}$ and conditional Gaussian distribution for $\mathbf{y}$ given $\mathbf{x}$ in the form:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, \Lambda^{-1}) \tag{A.11}$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|A\mathbf{x} + \mathbf{b}, L^{-1}) \tag{A.12}$$

The marginal distribution of $\mathbf{y}$ and the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ are given by:

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|A\mu + \mathbf{b}, L^{-1} + A\Lambda^{-1}A^T) \tag{A.13}$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\Sigma\{A^T L(\mathbf{y} - \mathbf{b}) + \Lambda\mu\}, \Sigma) \tag{A.14}$$

where $\Sigma = (\Lambda + A^T L A)^{-1}$ and $\Lambda \equiv \Sigma^{-1}$.

A.3 Student's t distribution

This is a Student's t distribution, which is generalised to a location-scale family, that are a location parameter μ and a scale parameter Λ, which the density is defined by:

$$\mathrm{St}(\mathbf{x}|\mu,\Lambda,\nu) = \frac{\Gamma(0.5(\nu+D))}{\Gamma(0.5\nu)}\frac{\Lambda^{1/2}}{(\nu\pi)^{D/2}}\left[1+\frac{(\mathbf{x}-\mu)^T\Lambda(\mathbf{x}-\mu)}{\nu}\right]^{-0.5(\nu+D)} \tag{A.15}$$

$$\mathbb{E}[\mathbf{x}] = \mu \tag{A.16}$$

$$\mathrm{cov}[\mathbf{x}] = \frac{\nu}{\nu-2}\Lambda^{-1} \quad \text{for} \quad \nu > 2 \tag{A.17}$$

where $\mathbb{E}[\mathbf{x}]$, $\mathrm{cov}[\mathbf{x}]$ and ν are mean, covariance and degree of freedom of a Student-t distribution.

A Student's t-distribution arises naturally in many Bayesian inference problems. If we have Gaussian distribution, $\mathcal{N}(\mathbf{x}|\mu,(\eta\Lambda)^{-1})$, together with a Gamma distribution, $\mathrm{Gam}(\eta|\nu/2,\nu/2)$, and the precision is integrated out, to obtain the marginal distribution of $\mathbf{x}$ in the form:

$$\begin{aligned}\mathrm{St}(\mathbf{x}|\mu,\Lambda,\nu) &= \int_0^\infty \mathcal{N}(\mathbf{x}|\mu,(\eta\Lambda)^{-1})\,\mathrm{Gam}(\eta|\nu/2,\nu/2)\,d\eta \\ &= \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\frac{|\Lambda|^{1/2}}{(2\pi)^{D/2}}\int_0^\infty \eta^{D/2}\eta^{v/2-1}\exp^{-\nu\eta/2}\exp^{-\eta\Delta^2/2}\end{aligned} \tag{A.18}$$

where Δ^2 is the squared Mahalanobis distance defined by:

$$\Delta^2 = (\mathbf{x}-\mu)^T\Lambda(\mathbf{x}-\mu) \tag{A.19}$$

Now, the change of variable can be made by:

$$\tau = \eta\left[\frac{\nu}{2}+\frac{1}{2}\Delta^2\right]^{-1} \tag{A.20}$$

which gives a Student-t distribution:

$$\begin{aligned}\mathrm{St}(\mathbf{x}|\mu,\Lambda,\nu) &= \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\frac{|\Lambda|^{1/2}}{(2\pi)^{D/2}}\left[\frac{\nu}{2}+\frac{1}{2}\Delta^2\right]^{-0.5(D+\nu)}\int_0^\infty \tau^{D/2+\nu/2-1}\exp^{-\tau}\,d\tau \\ &= \frac{\Gamma(0.5(\nu+D))}{\Gamma(0.5\nu)}\frac{\Lambda^{1/2}}{(\nu\pi)^{D/2}}\left[1+\frac{\Delta^2}{\nu}\right]^{-0.5(\nu+D)} \qquad \text{(A.21)}\end{aligned}$$

A.4 Wishart distribution

The Wishart distribution is the multidimensional generalisation of the Gamma distribution, defined by:

$$\mathcal{W}(\Lambda|W,\nu) = B(W,\nu)|\Lambda|^{(\nu-D-1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}(W^{-1}\Lambda)\right) \qquad \text{(A.22)}$$

where

$$B(W,\nu) \equiv |W|^{-\nu/2}\left(2^{\nu D/2}\pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1} \qquad \text{(A.23)}$$

$$\mathbb{E}[\Lambda] = \nu W \qquad \text{(A.24)}$$

$$\mathbb{E}[\ln|\Lambda|] = \sum_{i=1}^{D}\psi\left(\frac{\nu+1-i}{2}\right) + D\ln 2 + \ln|W| \qquad \text{(A.25)}$$

$$H[\Lambda] = -\ln B(W,\nu) - \frac{\nu-D-1}{2}\mathbb{E}[\ln|\Lambda|] + \frac{\nu D}{2} \qquad \text{(A.26)}$$

where W is a $D\times D$ symmetric, positive definite matrix and ν is the number of degree of freedom of the distribution, restricted to $\nu > D-1$ to ensure that the Gamma function in the normalisation factor is well defined.

Appendix B

Bayesian Regression 2

This appendix describes detailed derivation of posterior and predictive distributions, used in Bayesian Regression 2 (BR-2).

B.1 Posterior distribution

From BR-2 method, recall the prior distribution, equation (3.3), given by:

$$\begin{aligned} p(\mathbf{w}, \sigma^2) &= p(\mathbf{w}|\sigma^2)\, p(\sigma^2) \\ &= \mathcal{N}(\bar{\mathbf{w}}, \sigma^2 V)\text{IG}(a, b) \\ &= \text{NIG}(\bar{\mathbf{w}}, V, a, b) \\ &= \frac{b^a}{(a\pi)^{D/2}|V|^{1/2}\Gamma(a)}(\sigma^2)^{-(a+0.5D+1)} \exp\left\{ -\frac{(\mathbf{w}-\bar{\mathbf{w}})^T V^{-1}(\mathbf{w}-\bar{\mathbf{w}}) + 2b}{2\sigma^2} \right\} \end{aligned} \tag{B.1}$$

Recall the likelihood function, equation (3.6), given by:

$$\begin{aligned} p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) &= \mathcal{N}(\mathbf{\Phi w}, \sigma^2 I) \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{(\mathbf{y}-\mathbf{\Phi w})^T(\mathbf{y}-\mathbf{\Phi w})}{2\sigma^2} \right\} \end{aligned} \tag{B.2}$$

The posterior distribution can be computed using Bayes' theorem, which can be expressed mathematically as:

$$\begin{aligned} p(\mathbf{w}, \sigma^2|\mathbf{y}, \mathbf{X}) &\propto \underbrace{p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2)}_{\text{likelihood}} \underbrace{p(\mathbf{w}|\sigma^2)p(\sigma^2)}_{\text{prior}} \\ &= \mathcal{N}(\mathbf{\Phi w}, \sigma^2 I)\, \text{NIG}(\bar{\mathbf{w}}, V, a, b) \\ &= \frac{b^a}{(2\pi)^{0.5(N+D)}|V|^{1/2}\,\Gamma(a)}(\sigma^2)^{-(a+0.5(N+D)+1)} \\ &\quad \exp\left\{ -\frac{(\mathbf{w}-\bar{\mathbf{w}})^T V^{-1}(\mathbf{w}-\bar{\mathbf{w}}) + 2b + (\mathbf{y}-\mathbf{\Phi w})^T(\mathbf{y}-\mathbf{\Phi w})}{2\sigma^2} \right\} \end{aligned} \tag{B.3}$$

Owing to the conjugate choice, the posterior distribution is also obtained in the same form as the prior: a NIG distribution, defined by:

$$\begin{aligned} p(\mathbf{w}, \sigma^2) = &\frac{(b^*)^{a*}}{(2\pi)^{0.5D}|V^*|^{1/2}\,\Gamma(a^*)}(\sigma^2)^{-(a^*+0.5D+1)} \\ &\exp\left\{ -\frac{(\mathbf{w}-\mathbf{w}^*)(V^*)^{-1}(\mathbf{w}-\mathbf{w}^*) + 2b^*}{2\sigma^2} \right\} \end{aligned} \tag{B.4}$$

Therefore by using identity, equation (B.3) is equivalent to equation (B.4). Firstly, in normalisation part, it is easy to define that:

$$a^* = a + 0.5N \tag{B.5}$$

Next, in exponential part, this can be written as:

$$\begin{aligned} &(\mathbf{w}-\bar{\mathbf{w}})^T V^{-1}(\mathbf{w}-\bar{\mathbf{w}}) + 2b + (\mathbf{y}-\mathbf{\Phi w})^T(\mathbf{y}-\mathbf{\Phi w}) \\ &\quad \equiv (\mathbf{w}-\mathbf{w}^*)(V^*)^{-1}(\mathbf{w}-\mathbf{w}^*) + 2b^* \\ &\mathbf{w}^T(V^{-1}+\mathbf{\Phi}^T\mathbf{\Phi})\mathbf{w} - 2\mathbf{w}^T(\mathbf{\Phi}^T\mathbf{y} + V^{-1}\bar{\mathbf{w}}) + \bar{\mathbf{w}}^T V^{-1}\bar{\mathbf{w}} + \mathbf{y}^T\mathbf{y} + 2b \\ &\quad \equiv \mathbf{w}^T(V^*)^{-1}\mathbf{w} - 2\mathbf{w}^T(V^*)^{-1}\mathbf{w}^* + (\mathbf{w}^*)^T(V^*)^{-1}\mathbf{w}^* + 2b^* \end{aligned}$$

This results in:

$$V^* = (V^{-1} + \mathbf{\Phi}^T\mathbf{\Phi})^{-1} \tag{B.6}$$

and

$$\begin{aligned} -2\mathbf{w}^T(V^*)^{-1}\mathbf{w}^* &= -2\mathbf{w}^T(\mathbf{\Phi}^T\mathbf{y} + V^{-1}\bar{\mathbf{w}}) \\ \mathbf{w}^* &= V^*(\mathbf{\Phi}^T\mathbf{y} + V^{-1}\bar{\mathbf{w}}) \end{aligned} \tag{B.7}$$

and

$$\begin{aligned} 2b^* + (\mathbf{w}^*)^T(V^*)^{-1}\mathbf{w}^* &= 2b + \bar{\mathbf{w}}^T V^{-1}\bar{\mathbf{w}} + \mathbf{y}^T\mathbf{y} \\ b^* &= \frac{1}{2}\left\{\bar{\mathbf{w}}^T V^{-1}\bar{\mathbf{w}} + \mathbf{y}^T\mathbf{y} - (\mathbf{w}^*)^T(V^*)^{-1}\mathbf{w}^*\right\} \end{aligned} \tag{B.8}$$

This is a Normal-Inverse-Gamma distribution, NIG$(\mathbf{w}^*, V^*, a^*, b^*)$, described in equation (3.8).

B.2 Predictive distribution

Predictive distribution represents prediction of $\mathbf{y}^*$ for new values of $\mathbf{x}^*$, which can be obtained by marginalising posterior distribution, equation (3.8), so that:

$$\begin{aligned} p(\mathbf{y}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) &= \int\int p(\mathbf{y}^*|\mathbf{X}^*, \mathbf{w}, \sigma^2)\, p(\mathbf{w}, \sigma^2)\, d\mathbf{w}\, d\sigma^2 \\ &= \int\int \mathcal{N}(\mathbf{\Phi}^*\mathbf{w}, \sigma^2 I) \times \text{NIG}(\mathbf{w}^*, V^*, a^*, b^*)\, d\mathbf{w}\, d\sigma^2 \end{aligned} \tag{B.9}$$

Using Bayes' theorem for Gaussian-Gamma variables (in this case Inverse Gamma), discussed in appendix A.3, predictive distribution of BR-2 results in a Student-t distribution, given by:

$$\text{St}(\mathbf{\Phi}^*\mathbf{w}^*, b^*(I + \mathbf{\Phi}^* V^* \mathbf{\Phi}^{*T}), 2a^*) \tag{B.10}$$

where

$$\mathbb{E}[\mathbf{y}^*] = \mathbf{\Phi}^*\mathbf{w}^* \tag{B.11}$$

$$\text{cov}[\mathbf{y}^*] = \frac{b^*}{a^* - 1}(I + \mathbf{\Phi}^* V^* \mathbf{\Phi}^{*T}) \tag{B.12}$$

where $\{\mathbf{X}^*, \mathbf{y}^*\}$, $\mathbb{E}[\mathbf{y}^*]$, $\text{cov}[\mathbf{y}^*]$ and $\nu = 2a^*$ are the new input-output values, mean, covariance and number degree of freedom, respectively.

Appendix C

Bayesian Regression 3 using Gibbs sampler

This appendix describes detail derivation of some posterior conditionals, used in Bayesian Regression 3 (BR-3) based on Gibbs sampler.

C.1 Posterior conditional distributions

A joint posterior distribution for all the parameters of BR-3, equation (4.10), can be computed using Bayes's theorem by:

$$p(\mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2 | \mathbf{y}_j) \propto \left[\prod_{j=1}^{M} p(\mathbf{y}_j | \mathbf{X}_j, \mathbf{w}_j, \sigma_j^2) p(\mathbf{w}_j | \bar{\mathbf{w}}, V^{-1}) p(\sigma_j^2 | a_j, b_j) \right] p(\bar{\mathbf{w}} | \eta, C) p(V^{-1} | \rho, R) \tag{C.1}$$

where M is the total number of units.

In order to obtain a posterior distribution for each parameter, it is necessary to derive their posterior conditional distributions. Each complete posterior conditional is proportional to the aforementioned joint posterior, equation (C.1). Thus, all of the terms in the

product of equation (C.1) that do not involve in each posterior conditional are absorbed into the normalising constant of this conditional. The complete posterior conditionals will be discussed next.

Posterior conditional for $\mathbf{w}_j$

The first posterior conditional, equation (4.11), can be written as:

$$\begin{aligned} p(\mathbf{w}_j|\mathbf{X}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2, \mathbf{y}_j) &\propto p(\mathbf{y}_j|\mathbf{X}_j, \mathbf{w}_j, \sigma_j^2) p(\mathbf{w}_j|\bar{\mathbf{w}}, V) \\ &= \mathcal{N}(\mathbf{\Phi}_j \mathbf{w}_j, \sigma_j^2 I)\, \mathcal{N}(\bar{\mathbf{w}}, V) \end{aligned} \tag{C.2}$$

Bayes' theorem for Gaussian variables stated in appendix A.2 can be used, resulting in:

$$p(\mathbf{w}_j|\mathbf{X}_j, \bar{\mathbf{w}}, V^{-1}, \sigma_j^2, \mathbf{y}_j) = \mathcal{N}(d_{\mathbf{w}_j}, D_{\mathbf{w}_j}) \tag{C.3}$$

where

$$D_{\mathbf{w}_j} = \left(\mathbf{\Phi}_j^T \mathbf{\Phi}_j / \sigma_j^2 + V^{-1}\right)^{-1} \tag{C.4}$$

$$d_{\mathbf{w}_j} = D_{\mathbf{w}_j} \left(\mathbf{\Phi}_j^T \mathbf{y}_j / \sigma_j^2 + V^{-1} \bar{\mathbf{w}}\right) \tag{C.5}$$

Posterior conditional for $\bar{\mathbf{w}}$

The second hierarchical level, $p(\mathbf{w}_j) \sim \mathcal{N}(\bar{\mathbf{w}}, V)$, is assumed to be independent and identically distributed (iid), the equation $\prod_{j=1}^{M} p(\mathbf{w}_j|\bar{\mathbf{w}}, V^{-1})$ can be written as:

$$\tilde{\mathbf{w}} = \tilde{I}\bar{\mathbf{w}} + \tilde{u} \tag{C.6}$$

where $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1^T\, \tilde{\mathbf{w}}_2^T \cdots \tilde{\mathbf{w}}_M^T]^T$, $\tilde{u} = [u_1^T\, u_2^T \cdots u_M^T]^T$, $\tilde{I} = [I_2\, I_2 \cdots I_2]^T$ and $E(\tilde{u}\tilde{u}^T) = I_M \otimes V$. Therefore the posterior conditional for $\bar{\mathbf{w}}$, equation (4.14), is given by:

$$p(\bar{\mathbf{w}}|\mathbf{X}_j, \mathbf{w}_j, V^{-1}, \sigma_j^2, \mathbf{y}_j) \propto \left[\prod_{j=1}^{M} p(\mathbf{w}_j|\bar{\mathbf{w}}, V^{-1})\right] p(\bar{\mathbf{w}}|\eta, C)$$
$$= \mathcal{N}(\tilde{I}\bar{\mathbf{w}}, (I_M \otimes V))\, \mathcal{N}(\eta, C) \tag{C.7}$$

Because both likelihood and prior are Gaussian, the principle in appendix A.2 can be applied, resulting in:

$$p(\bar{\mathbf{w}}|\mathbf{X}_j, \mathbf{w}_j, V^{-1}, \sigma_j^2, \mathbf{y}_j) = \mathcal{N}(d_{\bar{\mathbf{w}}}, D_{\bar{\mathbf{w}}}) \tag{C.8}$$

where

$$\begin{aligned} D_{\bar{\mathbf{w}}} &= (\tilde{I}^T(I_M \otimes V^{-1})\tilde{I} + C^{-1})^{-1} \\ &= \left(MV^{-1} + C^{-1}\right)^{-1} \qquad \text{(C.9)} \\ d_{\bar{\mathbf{w}}} &= D_{\bar{\mathbf{w}}}\left(\tilde{I}^T(I_M \otimes V^{-1})\bar{\mathbf{w}} + C^{-1}\eta\right) \\ &= D_{\bar{\mathbf{w}}}\left(MV^{-1}\bar{\mathbf{w}} + C^{-1}\eta\right) \qquad \text{(C.10)} \end{aligned}$$

where $\bar{\mathbf{w}} = \left(\frac{1}{M}\sum_{j=1}^{M} \mathbf{w}_j\right)$.

Posterior conditional for σ_j^2

The posterior conditional for noise variance, equation (4.17), can be calculated by:

$$\begin{aligned}
p(\sigma_j^2|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \mathbf{y}_j) &\propto p(\mathbf{y}_j|\mathbf{X}_j, \mathbf{w}_j, \sigma_j^2)p(\sigma_j^2|a_j, b_j) \\
&\propto \left[\prod_{j=1}^{M} p(\mathbf{y}_j|\mathbf{X}_j, \mathbf{w}_j, \sigma_j^2)p(\sigma_j^2|a_j, b_j))\right] \\
&= \frac{1}{(\sigma_j^2)^{(N\times M)/2}} \exp\left(-\frac{1}{2\sigma_j^2}(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)\right) \\
&\quad \times \frac{1}{(\sigma_j^2)^{(a_j+1)}} \exp\left(-\frac{1}{b_j\sigma^2}\right) \\
&= \frac{1}{(\sigma_j^2)^{((N\times M)/2+a+1)}} \\
&\quad \times \exp\left(-\frac{1}{\sigma_j^2}\left[\frac{1}{2}(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j) + b_j^{-1}\right]\right)
\end{aligned}$$

giving:

$$p(\sigma_j^2|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, V^{-1}, \mathbf{y}_j) = \text{IG}\left(\frac{N\times M}{2} + a, \left[\frac{1}{2}(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j) + b_j^{-1}\right]^{-1}\right) \tag{C.11}$$

Posterior conditional for V^{-1}

The final posterior conditional, equation (4.20), can be computed by:

$$\begin{aligned}
p(V^{-1}|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, \sigma_i^2, \mathbf{y}_j) &\propto \left[\prod_{j=1}^{M} p(\mathbf{w}_j|\bar{\mathbf{w}}, V^{-1})\right] p(V^{-1}|\rho, R) \\
&\propto |V^{-1}|^{(M/2)} \exp\left(-\frac{1}{2}\sum_{j=1}^{M}(\mathbf{w}_j - \bar{\mathbf{w}})^T V^{-1}(\mathbf{w}_j - \bar{\mathbf{w}})\right) \\
&\quad \times |V^{-1}|^{(\rho-D-1)/2} \exp\left(-\frac{1}{2}\text{tr}(\rho R V^{-1})\right) \\
&= |V^{-1}|^{(M+\rho-D-1)/2} \exp\left(-\frac{1}{2}\sum_{j=1}^{M}\text{tr}(\mathbf{w}_j - \bar{\mathbf{w}})(\mathbf{w}_j - \bar{\mathbf{w}})^T V^{-1} + \text{tr}(\rho R V^{-1})\right) \\
&= |V^{-1}|^{(M+\rho-D-1)/2} \exp\left[-\frac{1}{2}\text{tr}\left(\left[\sum_{j=1}^{M}(\mathbf{w}_j - \bar{\mathbf{w}})(\mathbf{w}_j - \bar{\mathbf{w}})^T + \rho R\right] V^{-1}\right)\right]
\end{aligned}$$

giving:

$$p(V^{-1}|\mathbf{X}_j, \mathbf{w}_j, \bar{\mathbf{w}}, \sigma^2, \mathbf{y}_j) = \mathcal{W}\left(\left[\sum_{j=1}^{M}(\mathbf{w}_j - \bar{\mathbf{w}})(\mathbf{w}_j - \bar{\mathbf{w}})^T + \rho R\right]^{-1}, M + \rho\right) \tag{C.12}$$

Appendix D

Bayesian Regression 3 using Variational Bayes

This appendix describes detail derivation of variational inference, used in Bayesian Regression 3 (BR-3). Since this is a main contribution of the research, this appendix attempts to provides deep discussion about the concept. First, the principle of variational inference concept is discussed in detail. Next, variational posterior and predictive distributions are derived in detail. Finally, the variational lower bound is also described in detail.

D.1 Variational Bayesian principle

This part describes verification of two main equations in variational Bayesian principles, which are the log marginal probability decomposition, equation (5.1), and general solution for optimal factors of variational posterior distributions, equation (5.5).

The first part is for verification of the log marginal probability decomposition, mentioned in equation (5.1). Firstly, the product rule of probability is used to give:

$$p(Y, \theta) = p(\theta|Y)\, p(Y) \tag{D.1}$$

Next, substituting equation (D.1) into equation (5.2), to give:

$$\begin{aligned}
\mathcal{L}(q) &= \int q(\theta) \ln \left\{ \frac{p(Y,\theta)}{q(\theta)} \right\} d\theta \\
&= \int q(\theta) \ln \left\{ \frac{p(\theta|Y)\, p(Y)}{q(\theta)} \right\} d\theta \\
&= \int q(\theta) \left\{ \ln \left[p(\theta|Y) \right] + \ln \left[p(Y) \right] - \ln[q(\theta)] \right\} d\theta \\
&= \underbrace{\int q(\theta) \ln \left\{ \frac{p(\theta|Y)}{q(\theta)} \right\} d\theta}_{\text{-KL}(q||p)} + \ln \left[p(Y) \right] \underbrace{\int q(\theta)\, d\theta}_{\text{sum to 1}}
\end{aligned} \tag{D.2}$$

Finally, substitute again, equation (D.2) into equation (5.1), to prove that:

$$\begin{aligned}
\ln p(Y) &= \mathcal{L}(q) + \mathrm{KL}(q||p) \\
&= -\mathrm{KL}(q||p) + \ln p(Y) + \mathrm{KL}(q||p) \\
&= \ln p(Y)
\end{aligned} \tag{D.3}$$

Another proof is for equation (5.5), which is the general solution for variational inference. Factorised distribution, equation (5.4), can be simplified as:

$$q(\theta) = \prod_{g=1}^{G} q_g(\theta_g) = \prod_{g} q_g = q_h \prod_{g \neq h} q_g \tag{D.4}$$

First, we substitute the factorised distributions, equation (5.4) or equation (D.4), to lower

bound, equation (5.2), to give:

$$
\begin{aligned}
\mathcal{L}(q) &= \int q(\theta) \ln \left\{ \frac{p(Y,\theta)}{q(\theta)} \right\} d\theta \\
&= \int \prod_g q_g \{ \ln p(Y,\theta) - \sum_g \ln q_g \} d\theta \\
&= \iint q_h \prod_{g \neq h} q_g \{ \ln p(Y,\theta) \}\, d\theta_g\, d\theta_h - \iint q_h \prod_{g \neq h} q_g \{ \ln q_h \sum_{g \neq h} \ln q_g \} d\theta_g\, d\theta_h \\
&= \int q_h \underbrace{\left\{ \int \prod_{g \neq h} q_g \{ \ln p(Y,\theta) \}\, d\theta_g \right\}}_{\ln \tilde{p}(Y,\theta_h)} d\theta_h - \underbrace{\int q_h \ln q_h\, d\theta_h \int \prod_{g \neq h} q_g \sum_{g \neq h} \ln q_g\, d\theta_g}_{\int q_h \ln q_h\, d\theta_h + \text{const}} \\
&= \int q_h \ln \tilde{p}(Y,\theta_h)\, d\theta_h - \int q_h \ln q_h\, d\theta_h + \text{const} \qquad \text{(D.5)} \\
&= \int q_h \ln \left\{ \frac{\tilde{p}(Y,\theta_h)}{q_h(\theta_h)} \right\} d\theta_h \qquad \text{(D.6)} \\
&= -\mathrm{KL}(q||p) \qquad \text{(D.7)}
\end{aligned}
$$

This is a negative KL-divergence between $q_h(\theta_h)$ and $\tilde{p}(Y,\theta_h)$. Therefore, maximising equation (D.5) is equivalent to minimising the KL-divergence, stated in equation (D.7). The minimum occurs when $q_h(\theta_h) = \tilde{p}(Y,\theta_h)$. As a result, a general expression for optimal solution $q_h^*(\theta_h)$ can be written as:

$$
\begin{aligned}
\ln q_h^*(\theta_h) &= \ln \tilde{p}(Y,\theta_h) \\
&= \int \ln p(Y,\theta) \prod_{g \neq h} q_g\, d\theta_g \\
&= \mathbb{E}_{g \neq h}[\ln p(Y,\theta)] + \text{const} \qquad \text{(D.8)}
\end{aligned}
$$

where const is a constant and $\mathbb{E}_{g \neq h}$ is the expectation with respect to variables present in in the remaining factors $g \neq h$. In other word, for each factor, the log of the joint distribution is taken over all variables and then average with respect to those variables not in that factor.

D.2 Variational posterior distribution

This part describes detail derivations of optimal factors of posterior distributions, explained in section 5.2.5. Our approach to variational inference is based on a factorised approximation to the true posterior distribution.

According to equation (5.15), there are five approximated posterior distributions with 10 parameter. They are $q(\mathbf{w}_j), q(\bar{\mathbf{w}}, \Lambda), q(\beta_j)$ and $q(\alpha)$, where $q(\bar{\mathbf{w}}, \Lambda) = q(\bar{\mathbf{w}}|\Lambda)\, q(\Lambda)$. Recall a general variational inference solutions, equation (5.5):

$$\ln q_h^*(\theta_h) = \mathbb{E}_{g \neq h}[\ln p(\mathbf{y}_j, \theta)] + \text{const} \tag{D.9}$$

where const is a constant and $\mathbb{E}_{g \neq h}$ is the expectation with respect to variables present in in the remaining factors $g \neq h$. Recall the joint log probability of the model, mentioned in equation (5.17):

$$\begin{aligned}
\ln p(\mathbf{y}_j, \theta) = & -\frac{1}{2}\sum_{j=1}^{M}\{\beta_j(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)\} \\
& - \frac{1}{2}\sum_{j=1}^{M}\{\beta_j(\mathbf{w}_j - \bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j - \bar{\mathbf{w}})\} \\
& - \frac{1}{2}\alpha(\bar{\mathbf{w}} - \eta)^T\Lambda(\bar{\mathbf{w}} - \eta) + \frac{1}{2}(\rho + MD - 1)\ln|\Lambda| \\
& - \frac{1}{2}\text{tr}(R^{-1}\Lambda) \\
& + \sum_{j=1}^{M}[\{\frac{1}{2}(N_j + MD) + (a_j^{\beta} - 1)\}\ln(\beta_j) - b_j^{\beta}\beta_j] \\
& + (a^{\alpha} - 1 + \frac{1}{2}D)\ln(\alpha) - b^{\alpha}\alpha
\end{aligned} \tag{D.10}$$

For each factor, the log of the joint distribution, equation (D.10), is taken over all variables and then average with respect to those variables not in that factor, according to equation (D.9).

Estimation for posterior $p(\mathbf{w}_j)$

This can be obtained by using the result of equation (D.10) which depend on $\mathbf{w}_j$ and substitute to general variational inference solutions, equation (D.9), to give:

$$\begin{aligned}\ln q^*(\mathbf{w}_j) = &-\frac{1}{2}\mathbb{E}_{\bar{\mathbf{w}},\Lambda,\beta_j}\left[\sum_{j=1}^{M}\beta_j(\mathbf{w}_j-\bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})\right]\\ &-\frac{1}{2}\mathbb{E}_{\beta_j}\left[\sum_{j=1}^{M}\beta_j(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)\right]\end{aligned} \tag{D.11}$$

where

$$\begin{aligned}&\mathbb{E}_{\beta_j}\left[\sum_{j=1}^{M}\beta_j(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)\right]\\ &=\sum_{j=1}^{M}\mathbb{E}[\beta_j](\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j-\mathbf{\Phi}_j\mathbf{w}_j)\end{aligned} \tag{D.12}$$

and

$$\begin{aligned}&\mathbb{E}_{\bar{\mathbf{w}},\Lambda,\beta_j}\left[\sum_{j=1}^{M}\beta_j(\mathbf{w}_j-\bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})\right]\\ &=\int\int\int\sum_{j=1}^{M}\left[\beta_j\left[\mathrm{Tr}\{\Lambda((\mathbf{w}_j-\bar{\mathbf{w}})(\mathbf{w}_j-\bar{\mathbf{w}})^T)\}\right]\right]q^*(\bar{\mathbf{w}}|\Lambda)q^*(\Lambda)q^*(\beta_j)\,d\bar{\mathbf{w}}\,d\Lambda\,d\beta_j\\ &=\sum_{j=1}^{M}\mathbb{E}[\beta_j]\{(\mathbf{w}_j-d_{\bar{\mathbf{w}}})^T\mathbb{E}[\Lambda](\mathbf{w}_j-d_{\bar{\mathbf{w}}})\}+D\big(\mathbb{E}[\alpha]+\sum_{j=1}^{M}\mathbb{E}[\beta_j]\big)^{-1}\end{aligned} \tag{D.13}$$

If the approximated posterior $p(\mathbf{w}_j)$ is given by:

$$\begin{aligned} q^*(\mathbf{w}_j) &= \mathcal{N}(\mathbf{w}_j|d_{\mathbf{w}_j}, (\mathbb{E}[\beta_j])^{-1}D_{\mathbf{w}_j}) \\ \ln q^*(\mathbf{w}_j) &= -\frac{1}{2}\sum_{j=1}^{M}\mathbb{E}[\beta_j]\big\{(\mathbf{w}_j - d_{\mathbf{w}_j})^T D_{\mathbf{w}_j}^{-1}(\mathbf{w}_j - d_{\mathbf{w}_j})\big\} \\ &\quad + \frac{MD}{2}\sum_{j=1}^{M}\big\{\ln|\mathbb{E}[\beta_j]| + \ln|D_{\mathbf{w}_j}|\big\} \end{aligned} \tag{D.14}$$

Therefore by using identity, equation (D.11) is equivalent to equation (D.14), to give:

$$D_{\mathbf{w}_j}^{-1} = (\mathbf{\Phi}_j^T\mathbf{\Phi}_j + \mathbb{E}[\Lambda]) \tag{D.15}$$

$$d_{\mathbf{w}_j} = D_{\mathbf{w}_j}(\mathbf{\Phi}_j^T\mathbf{y}_j + \mathbb{E}[\Lambda]d_{\bar{\mathbf{w}}}) \tag{D.16}$$

Estimation for posterior $p(\bar{\mathbf{w}}, \Lambda)$

Variational posterior distribution $q^*(\bar{\mathbf{w}}, \Lambda)$ does not factorise into the product of the marginals, but by using the product rule, it can be decomposed as:

$$q^*(\bar{\mathbf{w}}, \Lambda) = q^*(\bar{\mathbf{w}}|\Lambda)\, q^*(\Lambda) \tag{D.17}$$

The result of equation (D.10) which depend on $\{\bar{\mathbf{w}}, \Lambda\}$ is used and substituted to general variational inference solutions, equation (D.9), to give:

$$\begin{aligned} \ln q^*(\bar{\mathbf{w}}, \Lambda) &= \frac{1}{2}(\rho + MD - 1)\ln|\Lambda| - \frac{1}{2}\mathbb{E}_{\mathbf{w}_j,\beta_j}\left[\sum_{j=1}^{M}[\beta_j(\mathbf{w}_j - \bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j - \bar{\mathbf{w}})]\right] \\ &\quad - \frac{1}{2}\mathbb{E}[\alpha](\bar{\mathbf{w}} - \eta)^T\Lambda(\bar{\mathbf{w}} - \eta) - \frac{1}{2}\text{tr}(R^{-1}\Lambda) \end{aligned} \tag{D.18}$$

where

$$\begin{aligned}
&\mathbb{E}_{\mathbf{w}_j,\beta_j}\Big[\sum_{j=1}^{M}[\beta_j(\mathbf{w}_j-\bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})]\Big] \\
&\quad=\iint\sum_{j=1}^{M}\beta_j\,\mathrm{tr}\{\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})(\mathbf{w}_j-\bar{\mathbf{w}})^T\}q^*(\mathbf{w}_j)q^*(\beta_j)\,d\mathbf{w}_j\,d\beta_j \\
&\quad=\sum_{j=1}^{M}\big\{\mathbb{E}[\beta_j](d_{\mathbf{w}_j}-\bar{\mathbf{w}})^T\Lambda(d_{\mathbf{w}_j}-\bar{\mathbf{w}})+\Lambda D_{\mathbf{w}_j}\big\} \qquad \text{(D.19)}
\end{aligned}$$

Therefore, it gives:

$$\begin{aligned}
\ln q^*(\bar{\mathbf{w}},\Lambda)=&\frac{1}{2}(\rho+MD-1)\ln|\Lambda|-\frac{1}{2}\mathbb{E}[\alpha](\bar{\mathbf{w}}-\eta)^T\Lambda(\bar{\mathbf{w}}-\eta)-\frac{1}{2}\mathrm{tr}(R^{-1}\Lambda) \\
&-\frac{1}{2}\sum_{j=1}^{M}\big\{\mathbb{E}[\beta_j](d_{\mathbf{w}_j}-\bar{\mathbf{w}})^T\Lambda(d_{\mathbf{w}_j}-\bar{\mathbf{w}})+\Lambda D_{\mathbf{w}_j}\big\} \qquad \text{(D.20)}
\end{aligned}$$

For $\ln q^*(\bar{\mathbf{w}}|\Lambda)$, it is necessary to consider only terms on the right hand side of equation (D.20) which depend on $\bar{\mathbf{w}}$, to give:

$$\begin{aligned}
\ln q^*(\bar{\mathbf{w}}|\Lambda)=&-\frac{1}{2}\mathbb{E}[\alpha](\bar{\mathbf{w}}-\eta)^T\Lambda(\bar{\mathbf{w}}-\eta) \\
&-\frac{1}{2}\sum_{j=1}^{M}\big\{\mathbb{E}[\beta_j](d_{\mathbf{w}_j}-\bar{\mathbf{w}})^T\Lambda(d_{\mathbf{w}_j}-\bar{\mathbf{w}})+\Lambda D_{\mathbf{w}_j}\big\} \qquad \text{(D.21)}
\end{aligned}$$

If the approximated posterior $p(\bar{\mathbf{w}}|\Lambda)$ is given by:

$$\begin{aligned}
q^*(\bar{\mathbf{w}}|\Lambda)&=\mathcal{N}(\bar{\mathbf{w}}|d_{\bar{\mathbf{w}}},D_{\bar{\mathbf{w}}}) \\
\ln q^*(\bar{\mathbf{w}}|\Lambda)&=-\frac{1}{2}(\bar{\mathbf{w}}-d_{\bar{\mathbf{w}}})^TD_{\bar{\mathbf{w}}}^{-1}(\bar{\mathbf{w}}-d_{\bar{\mathbf{w}}})-\frac{D}{2}\ln|D_{\bar{\mathbf{w}}}| \qquad \text{(D.22)}
\end{aligned}$$

Therefore by using identity, equation (D.21) is equivalent to equation (D.22), to give:

$$D_{\bar{\mathbf{w}}} = r\Lambda^{-1} \tag{D.23}$$

$$d_{\bar{\mathbf{w}}} = r(\sum_{j=1}^{M} \mathbb{E}[\beta_j] d_{\mathbf{w}_j} + \mathbb{E}[\alpha]\eta) \tag{D.24}$$

where $r = (\mathbb{E}[\alpha] + \sum_{j=1}^{M} \mathbb{E}[\beta_j])^{-1}$. Next, we determine the form of $q^*(\Lambda)$ by making use of the relation:

$$\begin{aligned}
\ln q^*(\Lambda) &= \ln q^*(\bar{\mathbf{w}}, \Lambda) - \ln q^*(\bar{\mathbf{w}}|\Lambda) \\
&= \frac{1}{2}(\rho + MD - 1)\ln|\Lambda| - \frac{1}{2}\mathbb{E}[\alpha](\bar{\mathbf{w}} - \eta)^T\Lambda(\bar{\mathbf{w}} - \eta) \\
&\quad - \frac{1}{2}\mathrm{tr}(R^{-1}\Lambda) - \frac{1}{2}\Big[\sum_{j=1}^{M} \mathbb{E}[\beta_j](d_{\mathbf{w}_j} - \bar{\mathbf{w}})^T\Lambda(d_{\mathbf{w}_j} - \bar{\mathbf{w}}) + \Lambda D_{\mathbf{w}_j}\Big] \\
&\quad - \Bigg[-\frac{1}{2}(\bar{\mathbf{w}} - d_{\bar{\mathbf{w}}})^T D_{\bar{\mathbf{w}}}^{-1}(\bar{\mathbf{w}} - d_{\bar{\mathbf{w}}}) - \frac{D}{2}\ln|D_{\bar{\mathbf{w}}}|\Bigg]
\end{aligned} \tag{D.25}$$

For $\ln q^*(\Lambda)$, note that the terms involving $\bar{\mathbf{w}}$ have cancelled out in equation (D.25). Therefore, $q^*(\Lambda)$ is independent of $\bar{\mathbf{w}}$, results in:

$$\begin{aligned}
\ln q^*(\Lambda) = \frac{1}{2}(\rho + MD - 1)\ln|\Lambda| - \frac{1}{2}\mathbb{E}[\alpha]\mathrm{tr}(\Lambda\eta\eta^T) - \frac{1}{2}\mathrm{tr}(R^{-1}\Lambda) \\
- \frac{1}{2}\Bigg[\sum_{j=1}^{M} \mathbb{E}[\beta_j]\,\mathrm{tr}(\Lambda d_{\mathbf{w}_j} d_{\mathbf{w}_j}^T) + \Lambda D_{\mathbf{w}_j}\Bigg] \\
+ \frac{1}{2}r^{-1}\mathrm{tr}(\Lambda d_{\bar{\mathbf{w}}} d_{\bar{\mathbf{w}}}^T) + \frac{D}{2}(\ln|r| - \ln|\Lambda|)
\end{aligned} \tag{D.26}$$

If the approximated posterior $p(\Lambda)$ is given by:

$$\begin{aligned}
q^*(\Lambda) &= \mathcal{W}(\Lambda|\hat{\rho}, \hat{R}) \\
\ln q^*(\Lambda) &= \frac{1}{2}(\hat{\rho} - D - 1)\ln|\Lambda| - \frac{1}{2}\mathrm{tr}(\hat{R}^{-1}\Lambda)
\end{aligned} \tag{D.27}$$

Therefore by using identity, equation (D.26) is equivalent to equation (D.27), to give:

$$\hat{\rho} = \rho + M \tag{D.28}$$

$$\hat{R}^{-1} = R^{-1} + \sum_{j=1}^{M}(D_{\mathbf{w}_j} + \mathbb{E}[\beta_j] d_{\mathbf{w}_j} d_{\mathbf{w}_j}^T) + \mathbb{E}[\alpha]\eta\eta^T - r^{-1} d_{\bar{\mathbf{w}}} d_{\bar{\mathbf{w}}}^T \tag{D.29}$$

Estimation for posterior $p(\beta_j)$

The result of equation (D.10) which depend on β_j is used and substituted to general variational inference solutions, equation (D.9), to give:

$$\begin{aligned}\ln q^*(\beta_j) = \sum_{j=1}^{M} \big[\{0.5(N_j + MD) + (a_j^\beta - 1)\}\ln|\beta_j|\big] \\ - \frac{1}{2}\mathbb{E}_{\mathbf{w}_j,\bar{\mathbf{w}},\Lambda}\Bigg[\sum_{j=1}^{M}\beta_j(\mathbf{w}_j - \bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j - \bar{\mathbf{w}}) \\ + \sum_{j=1}^{M}\beta_j(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)\Bigg]\end{aligned} \tag{D.30}$$

where

$$\begin{aligned}\mathbb{E}_{\mathbf{w}_j}\Bigg[\sum_{j=1}^{M}\beta_j(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)^T(\mathbf{y}_j - \mathbf{\Phi}_j\mathbf{w}_j)\Bigg] \\ = \sum_{j=1}^{M}\Big[\beta_j\{(\mathbf{y}_j - \mathbf{\Phi}_j d_{\mathbf{w}_j})^T(\mathbf{y}_j - \mathbf{\Phi}_j d_{\mathbf{w}_j})\}\Big] \\ + \sum_{j=1}^{M}\Big[\beta_j\{\text{tr}[\mathbf{\Phi}_j^T\mathbf{\Phi}_j(\mathbb{E}[\beta_j])^{-1}D_{\mathbf{w}_j}]\}\Big]\end{aligned} \tag{D.31}$$

and

$$\begin{aligned}
&\mathbb{E}_{\mathbf{w}_j,\bar{\mathbf{w}},\Lambda}\Big[\sum_{j=1}^{M}\beta_j(\mathbf{w}_j-\bar{\mathbf{w}})^T\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})\Big] \\
&\quad=\int\int\int\sum_{j=1}^{M}\beta_j\text{tr}[\Lambda(\mathbf{w}_j-\bar{\mathbf{w}})(\mathbf{w}_j-\bar{\mathbf{w}})^T]\,q^*(\mathbf{w}_j)q^*(\bar{\mathbf{w}})q^*(\Lambda)\,d\mathbf{w}_j\,d\bar{\mathbf{w}}\,d\Lambda \\
&\quad=\sum_{j=1}^{M}\beta_j\big[(d_{\mathbf{w}_j}-d_{\bar{\mathbf{w}}})^T\mathbb{E}[\Lambda](d_{\mathbf{w}_j}-d_{\bar{\mathbf{w}}})+\text{tr}((\mathbb{E}[\beta_j])^{-1}\Lambda D_{\mathbf{w}_j})+\Lambda D_{\bar{\mathbf{w}}}\big] \quad \text{(D.32)}
\end{aligned}$$

If the approximated posterior $p(\beta_j)$ is given by:

$$\begin{aligned}
q(\beta_j)&=\text{Gam}(\beta_j|\hat{a}_j^{\beta},\hat{b}_j^{\beta}) \\
\ln q(\beta_j)&=\sum_{j=1}^{M}\big[(\hat{a}_j^{\beta}-1)\ln|\beta_j|-\hat{b}_j^{\beta}\big] \quad \text{(D.33)}
\end{aligned}$$

Therefore by using identity, equation (D.30) is equivalent to equation (D.33), to give:

$$\hat{a}_j^{\beta}=0.5(N_j+MD)+a_j^{\beta} \quad \text{(D.34)}$$

$$\begin{aligned}
\hat{b}_j^{\beta}=b_j^{\beta}+0.5\sum_{j=1}^{M}\Big[&(d_{\mathbf{w}_j}-d_{\bar{\mathbf{w}}})^T\mathbb{E}[\Lambda](d_{\mathbf{w}_j}-d_{\bar{\mathbf{w}}}) \\
&+(\mathbf{y}_j-\mathbf{\Phi}_j\,d_{\mathbf{w}_j})^T(\mathbf{y}_j-\mathbf{\Phi}_j\,d_{\mathbf{w}_j}) \\
&+\mathbb{E}^{-1}[\beta_j]\text{tr}(\mathbf{\Phi}_j^T\mathbf{\Phi}_jD_{\mathbf{w}_j}) \\
&+\mathbb{E}^{-1}[\beta_j]\text{tr}(\Lambda D_{\mathbf{w}_j})+rD\Big] \quad \text{(D.35)}
\end{aligned}$$

Estimation for posterior $p(\alpha)$

The result of equation (D.10) which depend on α is used and substituted to general variational inference solutions, equation (D.9), to give:

$$\ln q^*(\alpha)=(a^{\alpha}-1+\frac{1}{2}D)\ln|\alpha|-b^{\alpha}\alpha-\frac{1}{2}\,\alpha\,\mathbb{E}_{\bar{\mathbf{w}},\Lambda}[(\bar{\mathbf{w}}-\eta)\Lambda(\bar{\mathbf{w}}-\eta)] \quad \text{(D.36)}$$

where

$$\begin{aligned}
&\mathbb{E}_{\bar{\mathbf{w}},\Lambda}[(\bar{\mathbf{w}}-\eta)^T\Lambda(\bar{\mathbf{w}}-\eta)] \\
&\quad = \iint \text{tr}\{\Lambda(\bar{\mathbf{w}}-\eta)(\bar{\mathbf{w}}-\eta)^T\} q^*(\bar{\mathbf{w}}|\Lambda) q^*(\Lambda)\, d\bar{\mathbf{w}}\, d\Lambda \\
&\quad = (d_{\bar{\mathbf{w}}}-\eta)^T\mathbb{E}[\Lambda](d_{\bar{\mathbf{w}}}-\eta) + \Lambda D_{\bar{\mathbf{w}}}
\end{aligned} \tag{D.37}$$

If the approximated posterior $p(\alpha)$ is given by:

$$\begin{aligned}
q^*(\alpha) &= \text{Gam}(\alpha|\hat{a}^\alpha, \hat{b}^\alpha) \\
\ln q^*(\alpha) &= (\hat{a}^\alpha - 1)\ln|\alpha| - (\hat{b}^\alpha\alpha)
\end{aligned} \tag{D.38}$$

Therefore, by using identity, equation (D.36) is equivalent to equation (D.38), to give:

$$\hat{a}^\alpha = a^\alpha + \frac{D}{2} \tag{D.39}$$

$$\hat{b}^\alpha = b^\alpha + \frac{1}{2}\{(d_{\bar{\mathbf{w}}}-\eta)^T\mathbb{E}[\Lambda](d_{\bar{\mathbf{w}}}-\eta) + rD\} \tag{D.40}$$

D.3 Predictive distribution

This part describes in detail the derivation of predictive distribution, explained in section 5.2.7. The posterior predictive density of the future of a degradation signal, $\mathbf{y}_j^*$, can be computed deterministically by marginalising the approximated posteriors $q(\mathbf{w}_j)$ and $q(\beta_j)$. This results in a predictive Student-t distribution, given by:

$$\begin{aligned} p(\mathbf{y}_j^*|\mathbf{X}_j^*) &= \iint p(\mathbf{y}_j^*|\mathbf{X}_j^*, \mathbf{w}_j, \beta_j) p(\mathbf{w}_j) p(\beta_j)\, d\mathbf{w}_j d\beta_j \\ &= \iint p(\mathbf{y}_j^*|\mathbf{X}_j^*, \mathbf{w}_j, \beta_j) q(\mathbf{w}_j) q(\beta_j)\, d\mathbf{w}_j d\beta_j \\ &= \iint \mathcal{N}(\mathbf{y}_j^*|\mathbf{\Phi}_j^* \mathbf{w}_j, \beta_j) \mathcal{N}(\mathbf{w}_j|d_{\mathbf{w}_j}, (\mathbb{E}[\beta_j])^{-1} D_{\mathbf{w}_j}) \text{Gam}(\beta_j|\hat{a}_j^\beta, \hat{b}_j^\beta)\, d\mathbf{w}_j d\beta_j \end{aligned} \tag{D.41}$$

Using Bayes' theorem for Gaussian variables, according to appendix A.2, the predictive distribution can be simplified as:

$$p(\mathbf{y}_j^*|\mathbf{X}_j^*) = \int \mathcal{N}(\mathbf{\Phi}_j^*\, d_{\mathbf{w}_j}, \beta_j^{-1}(I + \mathbf{\Phi}_j^* D_{\mathbf{w}_j} \mathbf{\Phi}_j^{*T}) \text{Gam}(\beta_j|\hat{a}_j^\beta, \hat{b}_j^\beta) d\beta_j \tag{D.42}$$

Next, use Bayes' theorem for Gaussian-Gamma variables, according to appendix A.3, the predictive distribution is in the form of Student-t distribution, given by:

$$\begin{aligned} p(\mathbf{y}_j^*|\mathbf{X}_j^*) &= \frac{\Gamma(0.5(\nu+1))}{\Gamma(0.5\nu)} \left(\frac{\mathbb{E}[\beta_j]}{\pi\nu}\right)^{1/2} \\ &\quad \left[1 + \frac{\mathbb{E}[\beta_j](\mathbf{y}_j^* - \mathbf{\Phi}_j^*\, d_{\mathbf{w}_j})^T S_0^{-1} (\mathbf{y}_j^* - \mathbf{\Phi}_j^*\, d_{\mathbf{w}_j})}{\nu}\right] \\ &= \text{St}\left(\mathbf{\Phi}_j^*\, d_{\mathbf{w}_j}, (\mathbb{E}[\beta_j])^{-1}(I + \mathbf{\Phi}_j^* D_{\mathbf{w}_j} \mathbf{\Phi}_j^{*T}), 2a_j^\beta\right) \end{aligned} \tag{D.43}$$

where $S_0 = (I + \mathbf{\Phi}_j^* \, D_{\mathbf{w}_j} \, \mathbf{\Phi}_j^{*T})$. Therefore, the properties of this distribution are:

$$\mathbb{E}[\mathbf{y}_j^*] = \mathbf{\Phi}_j^* \, d_{\mathbf{w}_j} \tag{D.44}$$

$$\text{cov}[\mathbf{y}_j^*] = \frac{b_j^\beta}{a_j^\beta - 1}(I + \mathbf{\Phi}_j^* \, D_{\mathbf{w}_j} \, \mathbf{\Phi}_j^{*T}) \tag{D.45}$$

$$\nu = 2a_j^\beta \tag{D.46}$$

where $\mathbb{E}[\mathbf{y}_j^*]$, $\text{cov}[\mathbf{y}_j^*]$ and ν are mean, covariance and degree of freedom of predictive student-t distribution, respectively.

D.4 Variational lower bound

This part describes in detail the derivation of variational lower bound, $\mathcal{L}(q)$, discussed in section 5.2.6. Variational lower bound is useful to monitor the bound during the re-estimation in order to test for convergence. Using equation (5.2), variational lower bound can be computed by:

$$\mathcal{L}(q) = \int q(\theta) \ln \left\{ \frac{P(Y, \theta)}{P(\theta)} \right\} d\theta \tag{D.47}$$

$$= \mathbb{E}[\ln p(Y, \theta)] - \mathbb{E}[\ln p(\theta)] \tag{D.48}$$

$$= \mathbb{E}[\ln p(\mathbf{y}_j, \mathbf{w}_j, \bar{\mathbf{w}}, \Lambda, \beta_j, \alpha)] - \mathbb{E}[\ln q(\mathbf{w}_j, \bar{\mathbf{w}}, \Lambda, \beta_j, \alpha)] \tag{D.49}$$

where $\mathbb{E}[\ln p(Y, \theta)]$ can be decomposed as:

$$\begin{aligned} \mathbb{E}[\ln p(Y, \theta)] = \mathbb{E}[&\ln p(\mathbf{y}_j | \mathbf{X}_j, \mathbf{w}_j, \beta_j) + \ln p(\mathbf{w}_j) + \ln p(\bar{\mathbf{w}} | \Lambda) \\ &+ \ln p(\Lambda) + \ln p(\beta_j) + \ln p(\alpha))] \end{aligned} \tag{D.50}$$

where

$$\begin{aligned}
\mathbb{E}_{\mathbf{w}_j,\beta_j}[\ln p(\mathbf{y}_j|\mathbf{X}_j,\mathbf{w}_j,\beta_j)] &= \sum_{j=1}^{M}\frac{N_j}{2}\mathbb{E}[\ln|\beta_j|] \\
&\quad - \frac{1}{2}\sum_{j=1}^{M}\mathbb{E}[\beta_j](\mathbf{y}_j - \mathbf{\Phi}_j\, d_{\mathbf{w}_j})^T(\mathbf{y}_j - \mathbf{\Phi}_j\, d_{\mathbf{w}_j}) \\
&\quad - \frac{1}{2}\sum_{j=1}^{M}\text{tr}(\mathbf{\Phi}_j^T\mathbf{\Phi}_j D_{\mathbf{w}_j}) \qquad \text{(D.51)} \\
\mathbb{E}_{\mathbf{w}_j,\bar{\mathbf{w}},\Lambda,\beta_j}[\ln p(\mathbf{w}_j)] &= \sum_{j=1}^{M}\frac{MD}{2}\mathbb{E}[\ln|\beta_j|] + \frac{MD}{2}\mathbb{E}[\ln|\Lambda|] \\
&\quad - \frac{1}{2}\sum_{j=1}^{M}\mathbb{E}[\beta_j]\{(d_{\mathbf{w}_j} - d_{\bar{\mathbf{w}}})^T\mathbb{E}[\Lambda](d_{\mathbf{w}_j} - d_{\bar{\mathbf{w}}}) + rD\} \\
&\quad - \frac{1}{2}\sum_{j=1}^{M}\text{tr}[\Lambda D_{\mathbf{w}_j}] \qquad \text{(D.52)} \\
\mathbb{E}_{\bar{\mathbf{w}},\Lambda,\alpha}[\ln p(\bar{\mathbf{w}}|\Lambda)] &= \frac{D}{2}(\mathbb{E}[\ln|\alpha|] + \mathbb{E}[\ln|\Lambda|]) \\
&\quad - \frac{1}{2}\mathbb{E}[\alpha]\{(d_{\bar{\mathbf{w}}} - \eta)^T\mathbb{E}[\Lambda](d_{\bar{\mathbf{w}}} - \eta) + rD\} \qquad \text{(D.53)} \\
\mathbb{E}_{\Lambda}[\ln p(\Lambda)] &= \frac{1}{2}(\rho - D - 1)\mathbb{E}[\ln|\Lambda|] - \frac{1}{2}\text{tr}(R^{-1}\mathbb{E}[\Lambda]) \qquad \text{(D.54)} \\
\mathbb{E}_{\beta_j}[\ln p(\beta_j)] &= \sum_{j=1}^{M}(a_j^{\beta} - 1)\mathbb{E}[\ln|\beta_j|] - \sum_{j=1}^{M} b_j^{\beta}\mathbb{E}[\beta_j] \qquad \text{(D.55)} \\
\mathbb{E}_{\alpha}[\ln p(\alpha)] &= (a^{\alpha} - 1)\mathbb{E}[\ln|\alpha|] - b^{\alpha}\mathbb{E}[\alpha] \qquad \text{(D.56)}
\end{aligned}$$

and $\mathbb{E}[\ln p(\theta)]$ is negative entropy of $q(\theta)$ distributions. The entropies of those distributions, $H(\theta)$, are given by:

$$H(\theta) = H(\mathbf{w}_j) + H(\bar{\mathbf{w}},\Lambda) + H(\beta_j) + H(\alpha) \qquad \text{(D.57)}$$

where $H(\bar{\mathbf{w}}, \Lambda) = H(\bar{\mathbf{w}}|\Lambda) + H(\Lambda)$. The detail of each entropy are given by:

$$H(\mathbf{w}_j) = \frac{1}{2}\sum_{j=1}^{M} \ln |D_{\mathbf{w}_j}| - \frac{1}{2}\sum_{j=1}^{M} \ln |\mathbb{E}[\beta_j]| + \frac{D}{2}(1 + \ln |2\pi|) \tag{D.58}$$

$$\begin{aligned} H(\bar{\mathbf{w}}, \Lambda) = &\frac{1}{2}\ln |D_{\bar{\mathbf{w}}}| + \frac{D}{2}(1 + \ln |2\pi|) \\ &- \ln B(\hat{R}, \hat{\rho}) - \frac{\hat{\rho} - D - 1}{2}\mathbb{E}[\ln |\Lambda|] + \frac{\hat{\rho} D}{2} \end{aligned} \tag{D.59}$$

$$H(\beta_j) = \sum_{j=1}^{M} \ln \Gamma(\hat{a}_j^{\beta}) - \sum_{j=1}^{M} (\hat{a}_j^{\beta} - 1)\psi(\hat{a}_j^{\beta}) - \sum_{j=1}^{M} \ln |\hat{b}_j^{\beta}| + \sum_{j=1}^{M} \hat{a}_j^{\beta} \tag{D.60}$$

$$H(\alpha) = \ln \Gamma(\hat{a}^{\alpha}) - (\hat{a}^{\alpha} - 1)\psi(\hat{a}^{\alpha}) - \ln \hat{b}^{\alpha} + \hat{a}^{\alpha} \tag{D.61}$$

where

$$\ln B(\hat{R}, \hat{\rho}) = \quad -\frac{\hat{\rho}}{2}\ln |\hat{R}| - \left\{ \frac{\hat{\rho} D}{2}\ln |2| + \frac{1}{4}D(D-1)\ln |\pi| + \sum_{i=1}^{D} \ln \Gamma\left(\frac{\hat{\rho} + 1 - i}{2}\right) \right\}$$

and $\psi(\hat{a}_j^{\beta})$ and $\psi(\hat{a}^{\alpha})$ are digamma functions. As a result, the simplified lower bound $\mathcal{L}(q)$ is given by:

$$\begin{aligned} \mathcal{L}(q) = &\frac{1}{2}(MD - M + D)\mathbb{E}[\ln|\Lambda|] - \frac{1}{2}\mathrm{tr}(R^{-1}\mathbb{E}[\Lambda]) \\ &- \hat{a}^{\alpha}\ln |\hat{b}^{\alpha}| - \sum_{j=1}^{M} \hat{a}_j^{\beta}\ln |\hat{b}_j^{\beta}| \\ &+ \ln \Gamma(\hat{a}^{\alpha}) + \sum_{j=1}^{M} \ln \Gamma(\hat{a}_j^{\beta}) \\ &+ \frac{1}{2}\sum_{j=1}^{M} \ln |D_{\mathbf{w}_j}| - \frac{1}{2}\sum_{j=1}^{M} \ln |\mathbb{E}[\beta_j]| + D(1 + \ln (2\pi)) + \frac{1}{2}\ln |D_{\bar{\mathbf{w}}}| \\ &+ \frac{\hat{\rho}}{2}\ln |\hat{R}| + \frac{\hat{\rho} D}{2}(1 + \ln |2|) + \frac{1}{4}D(D-1)\ln |\pi| \\ &+ \sum_{i=1}^{D} \ln \Gamma\left(\frac{\hat{\rho} + 1 - i}{2}\right) \end{aligned} \tag{D.62}$$

Appendix E

Comparison to Industrial Partner's Approach

E.1 Introduction

This appendix aims to provide algorithm analysis and several recommendations to practitioners. We emphasise the strength of our best approach, BR-3-VB with CPD (BR-3-VB+CPD), by comparing it to an industrial partner's approach; Linear Regression-OSyS (LR-OSyS) as well as a popular Bayesian approach, BR-1. LR-OSyS is a linear regression method based on a legacy approach developed in Optimized Systems & Solutions (OSyS)[1], whereas BR-1 has been discussed in section 2.4.2 and chapter 3, whilst BR-3-VB and CPD algorithm have been described in chapters 5 and 6, respectively.

Linear regression is used in industry due to its simplicity and computational speed. This approach is easy to implement, because prior knowledge is not required to run the algorithm, instead the algorithm relies solely on observed data. This method is also a deterministic method, which has a closed-form solution, such as least squares or maximum likelihood estimations for inferring model parameters.

[1]OSyS is a wholly-owned, independent subsidiary of the Rolls-Royce Group, providing a range of asset optimization products and managed services to a global customer base in the civil aviation, energy and power generation, oil and gas, civil nuclear, defence and marine markets.

In particular, LR-OSyS uses least squares method which requires 200 initial flight data points to ensure the robustness of the prediction, i.e. the prediction trajectory should be reasonable and guaranteed to extrapolate downward. Then, RUL estimation is updated every month. This method uses 200 window flight data points to capture variation in the rate of change of degradation. In this analysis, LS-OSyS is redeveloped. However, it does not consider the requirement of 200 initial flight data points and it updates every data point. The reasons are to make a fair comparison and because some degradation data have not reached 200 data points. Table E.1 summarises the implementation of these prognostic algorithms; LR, BR-1 and BR-3-VB+CPD.

E.2 Analysis

Using the same methodology as described in chapters 3-6, these three prognostic algorithms are implemented directly to operational or in-service data, the take-off values of TGT margin, generated from 60 gas turbine engines. The analysis is performed only for TGT margin degradation between two maintenance events.

Figure E.1 shows three examples of degradation data and RULs. Figures E.1a, E.1c and E.1e are degradation data for engines 3, 22 and 56, respectively. The first example illustrates a signal with a slope change in degradation whereas the last two examples show degradation signals with very high noise. Symbol "+" represents TGT margin data. Furthermore, figures E.1b, E.1d and E.1f demonstrate RUL metrics for engine 3, 22 and 56, respectively. The symbols $+$, $\triangle$ and $\bigcirc$ are the RUL estimates for LR-OSyS, BR-1 and BR-3-VB, respectively, where the solid line is "expected RUL".

These RUL metrics show clearly that LR-OSyS is inconsistent in early degradation signal because this approach does not have component of prior distribution like in Bayesian approach. In other words, this situation occurs because LR-OSyS does not hold prior knowledge about the system degradation. This is the main reason why in practice LR-OSyS starts the prediction after 200 data point, that is for generating appropriate results after receiving enough data points. On the other hand, BR-1 and BR-3-VB+CPD have reasonable starting predictions due to the presence of prior distribution. However, after receiving more data points, BR-1 starts to fluctuate largely from the true RUL because the high noise in degradation data. In contrast, our approach, BR-

3-VB+CPD, is able deal with these degradation signals. Nevertheless, all approaches converge to the "expected RUL" value once they receive enough data points.

Figure E.2 shows boxplots representing RUL residual across all engines. The results of LR-OSyS, BR-1 and BR-3-VB+CPD are shown on the top, middle and bottom subfigures, respectively. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers. At the beginning, LR-OSyS does not show large variation because there are incomplete information. Some of their RULs are unknown because the predictions are inconsistent and therefore they do not cross the threshold. However, the median of LR-OSyS is very large, indicate that the early predictions are mostly incorrect. On the other hand, BR-1 and BR-3-VB have better predictions in early predictions owing to prior distribution component in Bayesian methods. After receiving more data points, the median of LR-OSyS and BR-1 still deviate largely from zero, which indicate that most of their predictions diverge from the "expected RUL" values. In contrast, BR-3-VB+CPD has better RUL estimations in general. Despite there are some whiskers around box plots, but the median and the edges are still close to zero RUL residual.

The obtained results emphasise the superiority of the developed approach, BR-3-VB+CPD, compared to other popular methods, LR-OSyS and BR-1 as well as other developed approaches (see results in chapters 3-6).

Table E.1: The characteristics of the prognostic algorithms.

LR-OSyS	BR-1	BR-3-VB+CPD
uses linear regression to fit and extrapolate degradation signal.	uses Bayesian regression to fit and extrapolate degradation signal.	uses Bayesian hierarchical model to fit and extrapolate degradation signal.
requires 200 flight data points to initiate first prediction.	is capable to start prediction as soon as possible due to the availability of prior distribution.	is capable to start prediction as soon as possible due to the availability of prior distribution.
updates new prediction every month.	updates new prediction every flight cycle.	updates new prediction every flight cycle.
relies on single engine data.	uses all available engine data for estimating prior, then it relies on single engine data for individual prediction.	uses all available engine data for estimating prior as well as individual prediction.
uses 200 window flight data points to capture variation in the rate of change in degradation.	has uncertainty around prediction to capture uncertainty in degradation.	has uncertainty around prediction to capture uncertainty in degradation. is able to deal with slow decay and rapid decay in degradation owing to capabilities of sharing information and change point detection algorithm.

E.3 Recommendation for practitioners

Based on these results as well as previous results, delivered in chapters 3-6, we provide a strong recommendation to Rolls-Royce plc and OSyS to adopt and improvise our technique to be implemented in Enterprise Processing System (EPS)[1] for enhancing their existing fleet management.

Our developed prognostic approach is a generic algorithm which can be applied to various applications which are suitable to be dealt with data-driven framework. To complement the thesis discussion, we provide general recommendations to prognostic practitioners in implementing our prognostic algorithms to health monitoring systems, that are:

- Identify the physical phenomena associated with the evolution of the degradation process (crack, spalling and wear). This phenomena can be translated a ground truth, which is useful to validate RUL estimation. As described in this thesis, the absent of ground truth has proved a difficulty in validating the results.
- Choose the appropriate condition monitoring technology to monitor this phenomena, such as sensor system and data acquisition technologies.
- Identify a characteristic pattern in the sensory information to help develop a degradation model, e.g. linear and exponential growths. This pattern will be useful to assist users in determining a basis function of the prognostic model, such as first-order, second-order polynomials or other basis functions.
- Identify a failure threshold associated with the degradation signal.
- Implement prognostic algorithm based on our methodologies, shown in figures 3.1 or 4.1, depend upon the used algorithm. The outcome is RUL in the form of failure-time distribution. The maintenance decision can be made based on statistical mean, mode or the left tail of the failure-time distribution.

[1]Enterprise Processing System, or EPS, is the data analysis engine of the system in Optimized Systems & Solutions (OSyS). It manages data integration, storage, and orchestration of parallel data processing. Designed to be hugely scalable, it can address a global equipment base, or be deployed to a single machine.

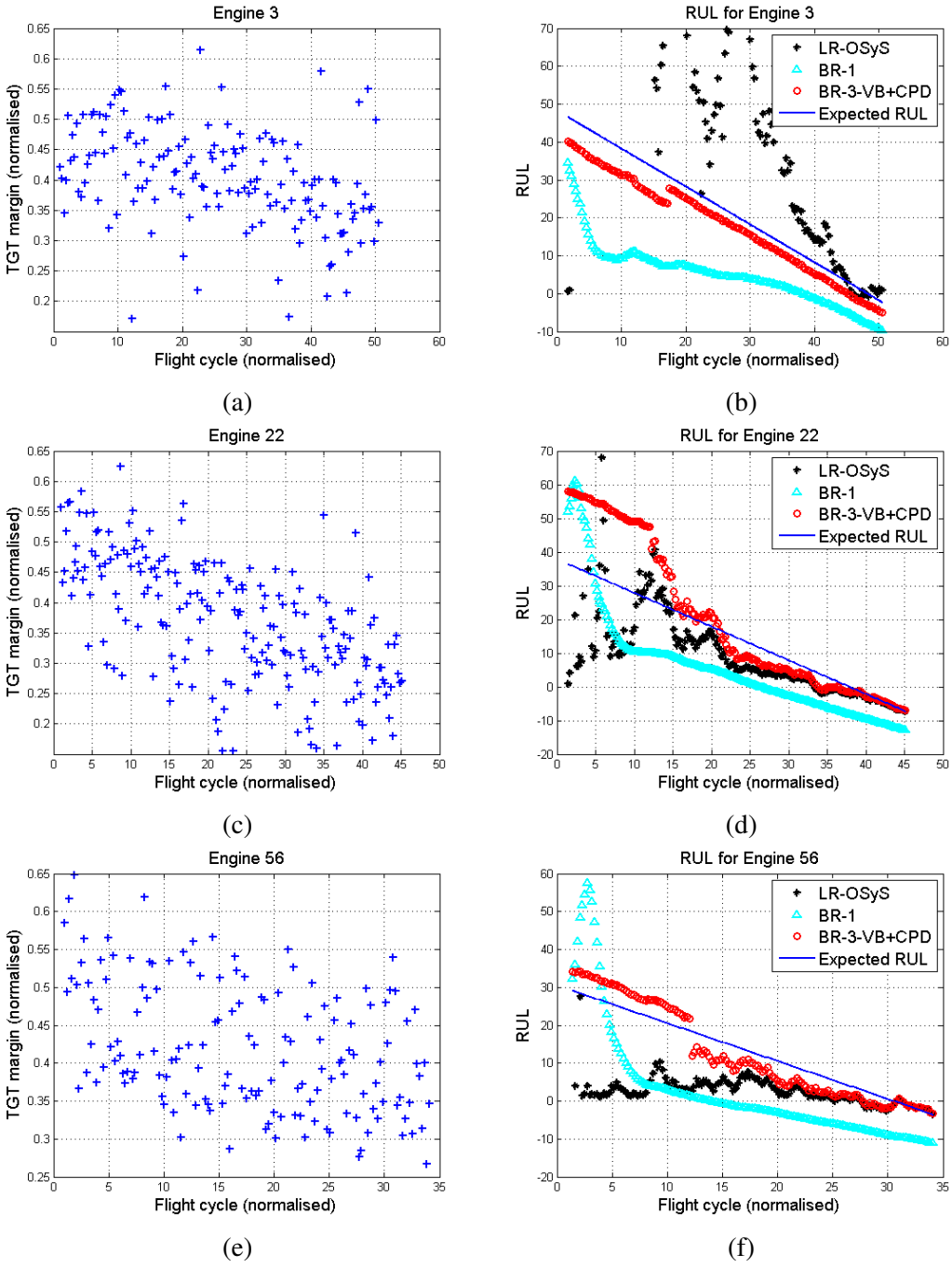


Figure E.1: Examples of TGT margin data and their RULs for engines 3, 22 and 56.

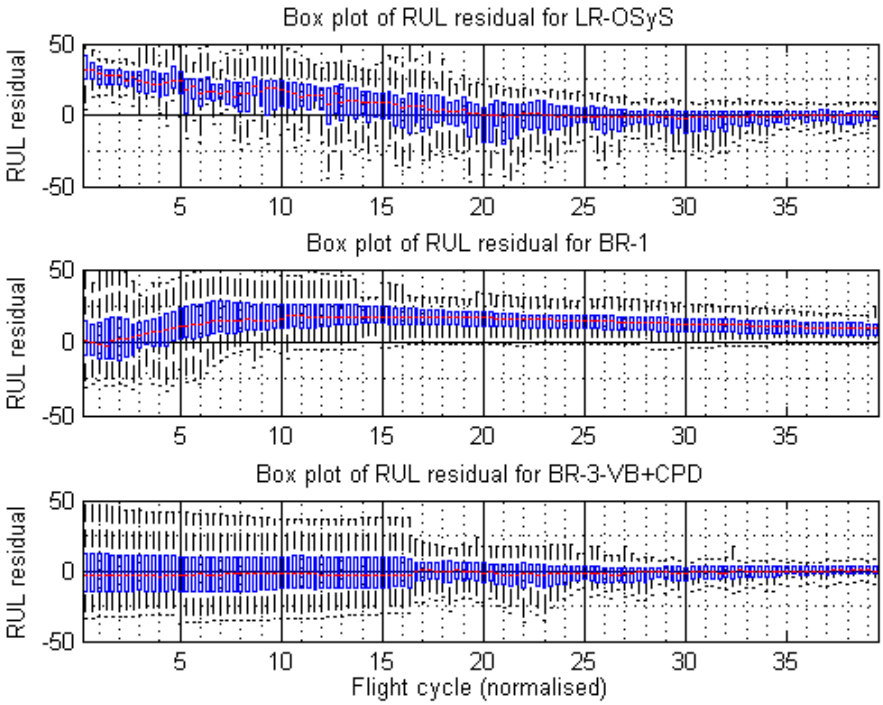


Figure E.2: An illustration of box plots of RUL residual across all 60 engines for LR-OSyS, BR-1 and BR-3-VB+CPD. The medians of LR-OSyS and BR-1 deviate largely from zero, which indicate that most of RUL estimations diverge from the "expected" RUL. It can be concluded that BR-3-VB+CPD has better RUL estimations than LR-OSyS and BR-1 methods.

References

Abbas, M., Ferri, A. A., Orchard, M. E., Vachtsevanos, G. J., 2007. An intelligent diagnostic/prognostic framework for automotive electrical systems. In: 2007 IEEE Intelligent Vehicles Symposium. IEEE, pp. 352–357.

Ackert, S., 2010. Engine maintenance concepts for financiers. Tech. rep., Aircraft Monitor.

Ahmad, M., Sheikh, A., 1984. Bernstein reliability model: Derivation and estimation of parameters. Reliability Engineering 8 (3), 131–148.

Alsyouf, I., 2004. Cost effective maintenance for competitive advantages. Ph.D. thesis, Vaxjo University.

Amemiya, T., 1994. Introduction to statistics and econometrics. Harvard University Press, Cambridge, Massachusetts.

Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical modeling and analysis for spatial data. Chapman & Hall/CRC, Boca Raton, Florida.

Banks, J., Merenich, J., 2007. Cost benefit analysis for asset health management technology. In: 2007 Proceedings - Annual Reliability and Maintainability Symposium. IEEE, pp. 95–100.

Banks, J., Reichard, K., Crow, E., Nickell, E., 2005. How engineers can conduct cost-benefit analysis for PHM systems. In: 2005 IEEE Aerospace Conference. IEEE, pp. 3958–3967.

Barber, D., 2012. Bayesian reasoning and machine learning. Cambridge University Press, New York.

Beal, M. J., 2003. Variational algorithms for approximate Bayesian inference. Ph.D. thesis, University of London.

Bernardo, J. M., Smith, A. F., 2009. Bayesian theory. John Wiley & Sons, New York.

Berry, D. A., 1996. Statistics: A Bayesian perspective. Duxbury Press, New York.

Biagetti, T., Sciubba, E., 2004. Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems. Energy 29 (12), 2553–2572.

Bishop, C. M., 2006. Pattern recognition and machine learning. Springer, New York.

Blei, D. M., Jordan, M. I., Ng, A. Y., 2003. Hierarchical Bayesian models for applications in information retrieval. In: Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., West, M. (Eds.), Bayesian Statistics 7. Oxford University Press, pp. 25–43.

Bond, L. J., Ramuhalli, P., Tawfik, M. S., Lybeck, N. J., 2011. Prognostics and life beyond 60 years for nuclear power plants. In: 2011 IEEE Conference on Prognostics and Health Management (PHM). IEEE, pp. 1–7.

Brasco, C., Eklund, N., Shah, M., Marthaler, D., 2013. Predictive modeling of high-bypass turbofan engine deterioration. In: 2013 Annual Conference of the Prognostics and Health Management Society. PHM Society, pp. 1–8.

Brodsky, B. E., Darkhovsky, B. S., 1993. Nonparametric methods in change point problems. Kluwer Academic Publishers Group, The Netherlands.

Brotherton, T., Jahns, G., Jacobs, J., Wroblewski, D., 2000. Prognosis of faults in gas turbine engines. In: 2000 IEEE Aerospace Conference. IEEE, pp. 163–171.

Brown, D. W., Georgoulas, G., Bole, B., Pei, H.-L., Orchard, M., Tang, L., Saha, B., Saxena, A., Goebel, K., Vachtsevanos, G., 2009. Prognostics enhanced reconfigurable control of electro-mechanical actuators. In: 2009 Annual Conference of the Prognostics and Health Management Society. PHM Society, pp. 1–17.

Brown, E., Moore, E., McCollom, N., Hess, A., 2007. Prognostics and health management a data-driven approach to supporting the F-35 lightning II. In: 2007 IEEE Aerospace Conference. IEEE, pp. 1–12.

Butler, K. L., 1996. An expert system based framework for an incipient failure detection and predictive maintenance system. In: 1996 International Conference on Intelligent Systems Applications to Power Systems, ISAP'96. IEEE, pp. 321–326.

Byington, C. S., Watson, M., Edwards, D., 2004a. Data-driven neural network methodology to remaining life predictions for aircraft actuator components. In: 2004 IEEE Aerospace Conference. IEEE, pp. 3581–3589.

Byington, C. S., Watson, M., Edwards, D., Stoelting, P., 2004b. A model-based approach to prognostics and health management for flight control actuators. In: 2004 IEEE Aerospace Conference. IEEE, pp. 3551–3562.

Caesarendra, W., Widodo, A., Yang, B.-S., 2010. Application of relevance vector machine and logistic regression for machine degradation assessment. Mechanical Systems and Signal Processing 24 (4), 1161–1171.

Canova, F., 2007. Methods for applied macroeconomic research. Princeton University Press, Princeton, New Jersey.

Canova, F., Ciccarelli, M., 2013. Panel vector autoregressive models: a survey. Tech. rep., European Central Bank.

Celaya, J., Saha, B., Wysocki, P., Goebel, K., 2008. Prognostics for electronics components of avionics systems. In: 2008 IEEE Aerospace Conference. IEEE, pp. 1–7.

Chakraborty, S., Gebraeel, N., Lawley, M., Wan, H., 2009. Residual-life estimation for components with non-symmetric priors. IIE Transactions 41 (4), 372–387.

Collett, D., 2003. Modelling survical data in medical research. Chapman & Hall/CRC, Boca Raton, Florida.

Cowles, M. K., Carlin, B. P., 1996. Markov chain monte carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association 91 (434), 883–904.

Csörgő, M., Horvath, L., 1988. Nonparametric methods for change-point problems. Handbook of statistics 7, 403–425.

Daniels, M. J., 1999. A prior for the variance in hierarchical models. Canadian Journal of Statistics 27 (3), 567–578.

Davey, N., 2003. The Gas Turbine-Development and Engineering. Watchmaker Publishing, Seaside, Oregon.

Davidian, M., Giltinan, D., 1995. Nonlinear models for repeated measurement data. Chapman & Hall/CRC, Boca Raton, Florida.

Del Negro, M., Schorfheide, F., 2011. Bayesian macroeconometrics. The Oxford Handbook of Bayesian Econometrics, 293–389.

Denison, D. G. T., Holmes, C. C., Mallick, B. K., Smith, A. F. M., 2002. Bayesian methods for nonlinear classification and regression. John Wiley & Sons, Chichester, UK.

Dennis, M., Kambil, A., 2003. Service management: Building profits after the sale. Supply Chain Management Review 7 (3), 42–48.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric reviews 3 (1), 1–100.

Drugowitsch, J., 2008. Bayesian linear regression. Tech. rep., University of Rochester.

Elwany, A., Gebraeel, N., 2009. Real-time estimation of mean remaining life using sensor-based degradation models. Journal of Manufacturing Science and Engineering 131 (5).

Eti, M. C., Ogaji, S., Probert, S., 2006. Development and implementation of preventive-maintenance practices in Nigerian industries. Applied energy 83 (10), 1163–1179.

Farrar, C. R., Cornwell, P. J., Hunter, N. F., Lieven, N. A., 2005. Sensing and data acquisition issues for damage prognosis. In: Damage Prognosis: For Aerospace, Civil and Mechanical Systems. John Wiley & Sons, Ltd, Chichester, UK, pp. 305–321.

Fei-Fei, L., Perona, P., 2005. A Bayesian hierarchical model for learning natural scene categories. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2005. IEEE, pp. 524–531.

Feldman, K., Jazouli, T., Sandborn, P., 2009. A methodology for determining the return on investment associated with prognostics and health management. IEEE Transactions on Reliability 58 (2), 305–316.

Feldman, K., Sandborn, P., 2008. Analyzing the return on investment associated with prognostics and health management of electronic products. In: 2008 ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE). ASME, pp. 1–9.

Feng, E., Yang, H., Rao, M., 1998. Fuzzy expert system for real-time process condition monitoring and incident prevention. Expert Systems with Applications 15 (3), 383–390.

Fox, C. W., Roberts, S. J., 2012. A tutorial on variational Bayesian inference. Artificial Intelligence Review 38 (2), 85–95.

Gao, Y., Li, Y., Zhang, G., Zhang, Z., 2008. Research status and perspectives of fault prediction technologies in prognostics and health management system. In: 2nd International Symposium on Systems and Control in Aerospace and Astronautics. ISSCAA 2008. IEEE, pp. 1–6.

Garg, S., 2004. Controls and health management technologies for intelligent aerospace propulsion systems. In: 42nd AIAA Aerospace Sciences Meeting and Exhibition. AIAA, pp. 1–24.

Garga, A., McClintic, K., Campbell, R., Yang, C.-C., Lebold, M., Hay, T., Byington, C., 2001. Hybrid reasoning for prognostic learning in cbm systems. In: 2001 IEEE Aerospace Conference. IEEE, pp. 2957–2969.

Gebraeel, N., Oct. 2006. Sensory-updated residual life distributions for components with exponential degradation patterns. IEEE Transactions on Automation Science and Engineering 3 (4), 382–393.

Gebraeel, N., Lawley, M., Li, R., Ryan, J., Jun. 2005. Residual-life distributions from component degradation signals: A Bayesian approach. IIE Transactions 37 (6), 543–557.

Gebraeel, N., Lawley, M., Liu, R., Parmeshwaran, V., 2004. Residual life predictions from vibration-based degradation signals: a neural network approach. IEEE Transactions on Industrial Electronics 51 (3), 694–700.

Gebraeel, N., Pan, J., 2008. Prognostic degradation models for computing and updating residual life distributions in a time-varying environment. IEEE Transactions on Reliability 57 (4), 539–550.

Gebraeel, N. Z., Lawley, M. A., 2008. A neural network degradation model for computing and updating residual life distributions. IEEE Transactions on Automation Science and Engineering 5 (1), 154–163.

Gelfand, A., Hills, S., Racine-Poon, A., Smith, A., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association 85 (412), 972–985.

Gelfand, A. E., Smith, A. F., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American statistical association 85 (410), 398–409.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). Bayesian Analysis 1 (3), 515–534.

Gelman, A., 2009. Bayes, Jeffreys, prior distributions and the philosophy of statistics. Statistical Science 24 (2), 176–178.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian data analysis, 2nd Edition. Chapman & Hall, Boca Raton, Florida.

Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York.

General-Electric, 2008. Understanding the EGT redline. GE Aviation Flight Operations Newsletter 3 (1), 4–7.

Giannone, D., Lenza, M., Primiceri, G. E., 2012. Prior selection for vector autoregressions. Tech. rep., European Central Bank.

Girolami, M., Rogers, S., 2005. Hierarchic bayesian models for kernel learning. In: Proceedings of the 22nd International Conference on Machine Learning. ICML '05. ACM, pp. 241–248.

Goebel, K., Bonissone, P., 2005. Prognostic information fusion for constant load systems. In: The 7th Annual Conference on Information Fusion. IEEE, pp. 1247–1255.

Goebel, K., Eklund, N., Bonanni, P., 2006. Fusing competing prediction algorithms for prognostics. In: 2006 IEEE Aerospace Conference. IEEE, pp. 1–10.

Goebel, K., Saha, B., Saxena, A., 2008a. A comparison of three data-driven techniques for prognostics. In: 62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT). MFPT, pp. 119–131.

Goebel, K., Saha, B., Saxena, A., Celaya, J., Christophersen, J., 2008b. Prognostics in battery health management. IEEE Instrumentation & Measurement Magazine 11 (4), 33–40.

Goh, K., Tjahjono, B., Baines, T., Subramaniam, S., 2006. A review of research in manufacturing prognostics. In: 2006 IEEE Conference on Industrial Informatics. IEEE, pp. 417–422.

Goodman, D. L., 2001. Prognostic methodology for deep submicron semiconductor failure modes. IEEE Transactions on Components and Packaging Technologies 24 (1), 109–111.

Graupe, D., 2007. Principles of artificial neural networks, 2nd Edition. Vol. 6. World Scientific, Singapore.

Hamada, M. S., Wilson, A. G., Reese, C. S., 2008. Bayesian reliability. Springer, New York.

Hecht, H., 2006. Why prognostics for avionics? In: 2006 IEEE Aerospace Conference. IEEE, pp. 1–6.

Heng, A., Zhang, S., Tan, A. C., Mathew, J., 2009. Rotating machinery prognostics: State of the art, challenges and opportunities. Mechanical Systems and Signal Processing 23 (3), 724–739.

Hill, P. G., Peterson, C. R., 1992. Mechanics and thermodynamics of propulsion. Addison-Wesley Publishing, Reading, Massachusetts.

Hjort, N. L., Holmes, C., Muller, P., Walker, S. G., 2010. Bayesian nonparametrics. Cambridge University Press, New York.

Huang, R., Xi, L., Li, X., Richard Liu, C., Qiu, H., Lee, J., 2007. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mechanical Systems and Signal Processing 21 (1), 193–207.

Ibrahim, J., Chen, M., Sinha, D., 2001. Bayesian survival analysis. Springer, New York.

Jaakkola, T. S., 2001. Tutorial on variational approximation methods. In: Opper, M., Saad, D. (Eds.), Advanced Mean Field Methods: Theory and Practice. The MIT Press, Cambridge, Massachusetts, pp. 129–160.

Jamshidi, M., 2010. Systems of systems engineering: principles and applications. CRC Press, Boca Raton, Florida.

Jardine, A. K., Lin, D., Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical systems and signal processing 20 (7), 1483–1510.

Jeffrey, H., 1961. Theory of probability. Clarendon Press, Oxford.

Johansson, J., Leisner, P., 2012. Prognostics of thermal fatigue failure of solder joints in avionic equipment. IEEE Aerospace and Electronic Systems Magazine 27 (4), 16–24.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., Saul, L. K., 1999. An introduction to variational methods for graphical models. Machine learning 37 (2), 183–233.

Kanamori, T., Hido, S., Sugiyama, M., 2009. A least-squares approach to direct importance estimation. The Journal of Machine Learning Research 10, 1391–1445.

Kass, R. E., Wasserman, L., 1996. The selection of prior distributions by formal rules. Journal of the American Statistical Association 91 (435), 1343–1370.

Katipamula, S., Brambley, M. R., 2005. Review article: Methods for fault detection, diagnostics, and prognostics for building systemsa review, part i. HVAC&R Research 11 (1), 3–25.

Kawahara, Y., Sugiyama, M., 2012. Sequential change-point detection based on direct density-ratio estimation. Statistical Analysis and Data Mining 5 (2), 114–127.

Kawahara, Y., Yairi, T., Machida, K., 2007. Change-point detection in time-series data based on subspace identification. In: Seventh IEEE International Conference on Data Mining. ICDM 2007. IEEE, pp. 559–564.

Khani, N., Segovia, C., Navaratne, R., Sethi, V., Singh, R., Pilidis, P., 2012. Towards development of a diagnostic and prognostic tool for civil aero-engine component degradation. In: 2012 ASME Gas Turbine India Conference. ASME, pp. 1–12.

Khawaja, T., Vachtsevanos, G., Wu, B., 2005. Reasoning about uncertainty in prognosis: a confidence prediction neural network approach. In: 2005 Annual Meeting of the North American on Fuzzy Information Processing Society. NAFIPS 2005. IEEE, pp. 7–12.

Kim, S.-H., Cohen, M. A., Netessine, S., 2007. Performance contracting in after-sales service supply chains. Management Science 53 (12), 1843–1858.

King, S., Bannister, P., Clifton, D., Tarassenko, L., 2009. Probabilistic approach to the condition monitoring of aerospace engines. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering 223 (5), 533–541.

Kleinbaum, D., Klein, M., 2005. Survival analysis: A self-learning text, 2nd Edition. Springer Verlag, New York.

Koop, G., Poirier, D., Tobias, L., 2007. Bayesian econometric methods (Econometric exercises). Cambridge University Press, New York.

Kothamasu, R., Huang, S. H., VerDuin, W. H., 2009. System health monitoring and prognostics–a review of current paradigms and practices. In: Handbook of Maintenance Management and Engineering. Springer, London, pp. 337–362.

Koudal, P., 2006. The service revolution in global manufacturing industries. Tech. rep., Deloitte Research.

Kullback, S., Leibler, R. A., 1951. On information and sufficiency. The Annals of Mathematical Statistics 22 (1), 79–86.

Kurz, R., Brun, K., 2001. Degradation in gas turbine systems. Journal of Engineering for Gas Turbines and Power (Transactions of the ASME) 123 (1), 70–77.

Kurz, R., Brun, K., 2007. Gas turbine tutorial maintenance and operating practices effects on degradation and life. In: The 36th Turbomachinery Symposium. ASME, pp. 173–186.

Leao, B., Fitzgibbon, K. T., Puttini, L. C., de Melo, G. P., 2008. Cost-benefit analysis methodology for PHM applied to legacy commercial aircraft. In: 2008 IEEE Aerospace Conference. IEEE, pp. 1–13.

Lebold, M., McClintic, K., Campbell, R., Byington, C., Maynard, K., 2000. Review of vibration analysis methods for gearbox diagnostics and prognostics. In: The 54th meeting of the society for Machinery Failure Prevention Technology. MFPT, pp. 623–634.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D., 2013. Prognostics and health management design for rotary machinery systems: Reviews, methodology and applications. Mechanical Systems and Signal Processing 42 (1), 314–334.

Lesaffre, E., Lawson, A. B., 2012. Bayesian biostatistics. John Wiley & Sons, Chichester, UK.

Li, C. J., Lee, H., 2005. Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. Mechanical systems and signal processing 19 (4), 836–846.

Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S., Liang, S., 1999. Adaptive prognostics for rolling element bearing condition. Mechanical systems and signal processing 13 (1), 103–113.

Li, Y., Nilkitsaranont, P., 2009. Gas turbine performance prognostic for condition-based maintenance. Applied Energy 86 (10), 2152–2161.

Lindley, D. V., Smith, A. F. M., 1972. Bayes estimates for the linear model. Journal of the Royal Statistical Society. Series B (Methodological) 34 (1), pp. 1–41.

Ling, Y., Mahadevan, S., 2012. Integration of structural health monitoring and fatigue damage prognosis. Mechanical Systems and Signal Processing 28, 89–104.

Lipowsky, H., Staudacher, S., Bauer, M., Schmidt, K. J., 2010. Application of Bayesian forecasting to change detection and prognosis of gas turbine performance. Journal of Engineering for Gas Turbines and Power 132 (3).

Liu, J., Wang, W., Ma, F., Yang, Y., Yang, C., 2012. A data-model-fusion prognostic framework for dynamic system state forecasting. Engineering Applications of Artificial Intelligence 25 (4), 814–823.

Liu, S., Yamada, M., Collier, N., Sugiyama, M., 2013. Change-point detection in time-series data by relative density-ratio estimation. Neural Networks 43 (0), 72 – 83.

Lu, C. J., Meeker, W. O., 1993. Using degradation measures to estimate a time-to-failure distribution. Technometrics 35 (2), 161–174.

Ma, Z., 2008. Survival analysis approach to reliability, survivability and prognostics and health management (PHM). In: 2008 IEEE Aerospace Conference. IEEE, pp. 1–20.

Ma, Z., Krings, A. W., 2008. Competing risks analysis of reliability, survivability, and prognostics and health management (PHM). In: 2008 IEEE Aerospace Conference. IEEE, pp. 1–21.

MacIsaac, B., Langton, R., 2011. Gas turbine propulsion systems. John Wiley & Sons, Chichester, UK.

Majidian, A., Saidi, M., 2007. Comparison of fuzzy logic and neural network in life prediction of boiler tubes. International Journal of Fatigue 29 (3), 489–498.

Makridakis, S., Wheelwright, S. C., Hyndman, R. J., 2008. Forecasting methods and applications. John Wiley & Sons, New York.

Malinge, Y., Courtenay, C., 2007. Avoiding high speed rejected takeoffs due to EGT limit exceedance. Safety First: the Airbus Safety Magazine (4), 8–13.

Marinai, L., 2004. Gas-path diagnostics and prognostics for aero-engines using fuzzy logic and time series analysis. Ph.D. thesis, Cranfield University.

Marinai, L., Probert, D., Singh, R., 2004. Prospects for aero gas-turbine diagnostics: a review. Applied Energy 79 (1), 109–126.

Marinai, L., Singh, R., Curnock, B., Probert, D., 2003. Detection and prediction of the performance deterioration of a turbofan engine. In: Proceedings of the International Gas Turbine Congress. GTSJ, pp. 1–7.

Marwala, T., Crossingham, B., 2008. Neuro-rough models for modelling hiv. In: 2008 IEEE Conference on Systems, Man and Cybernetics. SMC 2008. IEEE, pp. 3089–3095.

Menzefricke, U., 1999. Bayesian prediction in growth-curve models with correlated errors. Test 8 (1), 75–93.

Menzefricke, U., 2000. Hierarchical modeling with gaussian processes. Communications in Statistics-Simulation and Computation 29 (4), 1089–1108.

Minka, T. P., 2001a. A family of algorithms for approximate Bayesian inference. Ph.D. thesis, Massachusetts Institute of Technology.

Minka, T. P., 2001b. Expectation propagation for approximate Bayesian inference. In: Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., pp. 362–369.

Mobley, R. K., 2004. Maintenance fundamentals, 2nd Edition. Elsevier Butterworth-Heinemann, Burlington, Massachusetts.

Mohanty, S., Das, S., Chattopadhyay, A., Peralta, P., 2009. Gaussian process time series model for life prognosis of metallic structures. Journal of Intelligent Material Systems and Structures 20 (8), 887–896.

Moskvina, V., Zhigljavsky, A., 2003. An algorithm based on singular spectrum analysis for change-point detection. Communications in Statistics-Simulation and Computation 32 (2), 319–352.

Moubray, J., 1997. RCM II: reliability-centered maintenance, 2nd Edition. Industrial Press Inc., New York.

Müller, M., Staudacher, S., Friedl, W.-H., Köhler, R., Weißschuh, M., 2010. Probabilistic engine maintenance modeling for varying environmental and operating conditions. In: 2010 Proceedings of ASME Turbo Expo: Power for Land, Sea and Air. ASME, pp. 1–10.

Murphy, K. P., 2007. Conjugate Bayesian analysis of the Gaussian distribution. Tech. rep., University of British Columbia.

Nachlas, J. A., 2005. Reliability engineering: probabilistic models and maintenance methods. CRC Press, Boca Raton, Florida.

Naeem, M., Singh, R., Probert, D., 1998. Implications of engine deterioration for fuel usage. Applied energy 59 (2), 125–146.

Naeem, M., Singh, R., Probert, D., 2001. Consequences of aero-engine deteriorations for military aircraft. Applied energy 70 (2), 103–133.

Nusser, S., 2009. Robust learning in safety related domains: machine learning methods for solving safety related application problems. Ph.D. thesis, University Magdeburg.

O'Hagan, A., Luce, B., 2003. A primer on bayesian statistics in health economics and outcomes research. Tech. rep., the Centre for Bayesian Statistics in Health Economics, the University of Sheffield.

O'Hagan, A., Oakley, J. E., 2004. Probability is perfect, but we can't elicit it perfectly. Reliability Engineering & System Safety 85 (1), 239–248.

Ompusunggu, A. P., Vandenplas, S., Sas, P., Van Brussel, H., 2012. Health assessment and prognostics of automotive clutches. In: 2012 The First European Conference of the Prognostics and Health Management Society. PHM Society, pp. 114–127.

OSyS, 2013. OSyS Company Information. Accessed: Sept. 2013.
URL http://www.o-sys.com/about-us/our-company/

Panesar, S. S., Markeset, T., 2008. Development of a framework for industrial service innovation management and coordination. Journal of Quality in Maintenance Engineering 14 (2), 177–193.

Patil, N., Celaya, J., Das, D., Goebel, K., Pecht, M., 2009. Precursor parameter identification for insulated gate bipolar transistor (IGBT) prognostics. IEEE Transactions on Reliability 58 (2), 271–276.

Pecht, M., 2008. Prognostics and health management of electronics. John Wiley & Sons, Hoboken, New Jersey.

Pecht, M., Gu, J., 2009. Physics-of-failure-based prognostics for electronic products. Transactions of the Institute of Measurement and Control 31 (3-4), 309–322.

Pecht, M., Jaai, R., 2010. A prognostics and health management roadmap for information and electronics-rich systems. Microelectronics Reliability 50 (3), 317–323.

Peng, Y., Dong, M., Zuo, M. J., 2010. Current status of machine prognostics in condition-based maintenance: a review. The International Journal of Advanced Manufacturing Technology 50 (1-4), 297–313.

Pole, A., West, M., Harrison, J., 1994. Applied Bayesian forecasting and times series analysis. Chapman & Hall/CRC, Boca Raton, Florida.

Powrie, H., Fisher, C., 1999. Engine health monitoring: towards total prognostics. In: 1999 IEEE Aerospace Conference. Vol. 3. IEEE, pp. 11–20.

Przytula, K. W., Choi, A., 2008. An implementation of prognosis with dynamic Bayesian networks. In: 2008 IEEE Aerospace Conference. IEEE, pp. 1–8.

Puggina, N., Venturini, M., 2012. Development of a statistical methodology for gas turbine prognostics. Transactions of the ASME-A-Engineering for Gas Turbines and Power 134 (2), 1–9.

Qiu, H., Lee, J., Lin, J., Yu, G., 2003. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. Advanced Engineering Informatics 17 (3), 127–140.

Qiu, J., Seth, B. B., Liang, S. Y., Zhang, C., 2002. Damage mechanics approach for bearing lifetime prognostics. Mechanical systems and signal processing 16 (5), 817–829.

Ramakrishnan, A., Pecht, M. G., 2003. A life consumption monitoring methodology for electronic systems. IEEE Transactions on Components and Packaging Technologies 26 (3), 625–634.

Rasmussen, C. E., Williams, C. K. I., 2006. Gaussian processes for machine learning. The MIT press, Cambridge, Massachusetts.

Ray, A., Tangirala, S., 1996. Stochastic modeling of fatigue crack dynamics for on-line failure prognostics. IEEE Transactions on Control Systems Technology 4 (4), 443–451.

Rezaei, A., Dadouche, A., 2012. Development of a turbojet engine gearbox test rig for prognostics and health management. Mechanical Systems and Signal Processing 33 (0), 299 – 311.

Robinson, M., Crowder, M., 2000. Bayesian methods for a growth-curve degradation model with repeated measures. Lifetime Data Analysis 6 (4), 357–374.

Rolls-Royce, 1996. The jet engine, 5th Edition. Rolls-Royce plc, Derby, UK.

Rolls-Royce, 2013. Rolls-Royce: Our Corporate Story. Accessed: Sept. 2013. URL `http://www.rolls-royce.com/about/index.jsp`

Rossi, P. E., Allenby, G. M., McCulloch, R., 2005. Bayesian statistics and marketing. John Wiley & Sons, Chichester, UK.

Saha, B., Goebel, K., 2008. Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques. In: 2008 IEEE Aerospace Conference. IEEE, pp. 1–8.

Saha, B., Goebel, K., Poll, S., Christophersen, J., 2007. An integrated approach to battery health monitoring using bayesian regression and state estimation. In: 2007 IEEE Autotestcon. IEEE, pp. 646–653.

Saravanamuttoo, H., Rogers, G. F., Cohen, H., Straznicky, P., 2009. Gas turbine theory, sixth Edition. Pearson Education, Dorset, UK.

Satish, B., Sarma, N., 2005. A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors. In: 2005 IEEE Power Engineering Society General Meeting. IEEE, pp. 2291–2294.

Saunders, S. C., 2007. Reliability, Life Testing and the Prediction of Service Lives: For Engineers and Scientists. Springer, New York.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., Schwabacher, M., 2008a. Metrics for evaluating performance of prognostic techniques. In: 2008 IEEE Conference on Prognostics and Health Management. IEEE, pp. 1–17.

Saxena, A., Goebel, K., Simon, D., Eklund, N., 2008b. Damage propagation modeling for aircraft engine run-to-failure simulation. In: 2008 International Conference on Prognostics and Health Management. PHM 2008. IEEE, pp. 1–9.

Schwabacher, M., Goebel, K., 2007. A survey of artificial intelligence for prognostics. In: 2007 The Association for the Advancement of Artificial Intelligence Fall Symposium. AAAI, pp. 107–114.

Shao, Y., Nezu, K., 2000. Prognosis of remaining bearing life using neural networks. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 214 (3), 217–230.

Shaw, W. T., 2006. New methods for simulating the student t-distribution-direct use of the inverse cumulative distribution. Journal of Computational Finance 9 (4), 37–73.

Shenoy, D., Bhadury, B., et al., 1997. Maintenance Resource Management: Adapting Materials Requirements Planning (MRP). Taylor & Francis, Inc., London.

Si, X.-S., Wang, W., Hu, C.-H., Zhou, D.-H., 2011. Remaining useful life estimation–a review on the statistical data driven approaches. European Journal of Operational Research 213 (1), 1–14.

Sikorska, J., Hodkiewicz, M., Ma, L., 2011. Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing 25 (5), 1803–1836.

Sims, C. A., Zha, T., 1998. Bayesian methods for dynamic multivariate models. International Economic Review, 949–968.

Skaf, Z., Zaidan, M., Harrison, R., Mills, A., 2013. Accommodating repair actions into gas turbine prognostics. In: 2013 Annual conference of the Prognostics and Health Management society 2013. PHM Society, pp. 1–8.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. Bugs 0.5: Bayesian inference using gibbs sampling manual (version ii). MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, 1–59.

Sun, B., Zeng, S., Kang, R., Pecht, M. G., 2012. Benefits and challenges of system prognostics. IEEE Transactions on Reliability 61 (2), 323–335.

Tang, L., Kacprzynski, G. J., Goebel, K., Saxena, A., Saha, B., Vachtsevanos, G., 2008. Prognostics-enhanced automated contingency management for advanced autonomous systems. In: 2008 IEEE Conference on Prognostics and Health Management. PHM 2008. IEEE, pp. 1–9.

Tobias, J., 2001. Forecasting output growth rates and median output growth rates: A hierarchical bayesian approach. Journal of Forecasting 20 (5), 297–314.

Tran, V. T., Yang, B.-S., Oh, M.-S., Tan, A. C. C., 2008. Machine condition prognosis based on regression trees and one-step-ahead prediction. Mechanical Systems and Signal Processing 22 (5), 1179–1193.

Vapnik, V. N., 1998. Statistical learning theory. John Wiley & Sons, New York.

Walker, M., 2010. Next generation prognostics and health management for unmanned aircraft. In: 2010 IEEE Aerospace Conference. IEEE, pp. 1–14.

Walsh, P. P., Fletcher, P., 2004. Gas turbine performance. Blackwell Science Ltd, Oxford.

Wang, W., Pecht, M., 2011. Economic analysis of canary-based prognostics and health management. IEEE Transactions on Industrial Electronics 58 (7), 3077–3089.

Waters, N., Jun. 2009. Engine Health Management. Ingenia (39), 37–42.

West, M., Harrison, J., 1997. Bayesian forecasting and dynamic models. Springer, New York.

Widodo, A., Yang, B.-S., 2011a. Application of relevance vector machine and survival probability to machine degradation assessment. Expert Systems with Applications 38 (3), 2592–2599.

Widodo, A., Yang, B.-S., 2011b. Machine health prognostics using survival probability and support vector machine. Expert Systems with Applications 38 (7), 8430–8437.

Winn, J. M., Bishop, C. M., 2005. Variational message passing. In: Journal of Machine Learning Research. pp. 661–694.

Wireman, T., 2010. Training programs for maintenance organizations. Industrial Press Inc., New York.

Wu, W., Hu, J., Zhang, J., 2007. Prognostics of machine health condition using an improved arima-based prediction method. In: 2nd IEEE Conference on Industrial Electronics and Applications. ICIEA 2007. IEEE, pp. 1062–1067.

Xu, J., Xu, L., 2011. Health management based on fusion prognostics for avionics systems. Journal of Systems Engineering and Electronics 22 (3), 428–436.

Yam, R., Tse, P., Li, L., Tu, P., 2001. Intelligent predictive decision support system for condition-based maintenance. The International Journal of Advanced Manufacturing Technology 17 (5), 383–391.

Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H., Sugiyama, M., 2013. Relative density-ratio estimation for robust distribution comparison. Neural computation 25 (5), 1324–1370.

Yan, J., Koc, M., Lee, J., 2004. A prognostic algorithm for machine performance assessment and its application. Production Planning & Control 15 (8), 796–801.

Yu, M., Wang, D., Luo, M., Zhang, D., Chen, Q., 2012. Fault detection, isolation and identification for hybrid systems with unknown mode changes and fault patterns. Expert Systems with Applications 39 (11), 9955–9965.

Zanardelli, W. G., Strangas, E. G., Khalil, H. K., Miller, J. M., 2005. Wavelet-based methods for the prognosis of mechanical and electrical failures in electric motors. Mechanical Systems and Signal Processing 19 (2), 411–426.

Zhan, S., Rodiek, J., Heuermann-Kuehn, L. E., Baumann, J., 2011. Prognostics health management for a directional drilling system. In: 2011 Conference on Prognostics and Health Management. IEEE, pp. 1–7.

Zhang, L., Li, X., Yu, J., 2006. A review of fault prognostics in condition based maintenance. In: Sixth International Symposium on Instrumentation and Control Technology: Signal Analysis, Measurement Theory, Photo-Electronic Technology, and Artificial Intelligence. SPIE, pp. 1–8.

Zhang, X., Xu, R., Kwan, C., Liang, S. Y., Xie, Q., Haynes, L., 2005. An integrated approach to bearing fault diagnostics and prognostics. In: 2005 Proceedings of the American Control Conference. IEEE, pp. 2750–2755.