Unsupervised Methods for Condition-Based Maintenance in Non-Stationary Operating Conditions

by

Stanley Fong

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Wilson Wang Professor Dept. of Mechanical Engineering Lakehead University
Supervisor:	Sriram Narasimhan Professor Department of Civil and Environmental Engineering University of California, Los Angeles
Internal Members:	Giovanni Cascante Professor Department of Civil and Environmental Engineering University of Waterloo
	Chris Bachmann Assistant Professor Department of Civil and Environmental Engineering University of Waterloo
Internal-External Member:	Kaan Erkorkmaz Professor Department of Mechanical and Mechatronics Engineering University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Maintenance and operation of modern dynamic engineering systems requires the use of robust maintenance strategies that are reliable under uncertainty. One such strategy is condition-based maintenance (CBM), in which maintenance actions are determined based on the current health of the system. The CBM framework integrates fault detection and forecasting in the form of degradation modeling to provide real-time reliability, as well as valuable insight towards the future health of the system. Coupled with a modern information platform such as Internet-of-Things (IoT), CBM can deliver these critical functionalities at scale.

The increasingly complex design and operation of engineering systems has introduced novel problems to CBM. Characteristics of these systems—such as the unavailability of historical data, and highly dynamic operating behaviour of these systems—has rendered many existing solutions infeasible. These problems have motivated the development of new and self-sufficient—or in other words—unsupervised CBM solutions. The issue, however, is that many of the necessary methods required by such frameworks have yet to be proposed within the literature. Key gaps pertaining to the lack of suitable unsupervised approaches for the pre-processing of non-stationary vibration signals, parameter estimation for fault detection, and degradation threshold estimation, need to be addressed in order to achieve an effective implementation.

The main objective of this thesis is to propose set of three novel approaches to address each of the aforementioned knowledge gaps. A non-parametric pre-processing and spectral analysis approach, termed spectral mean shift clustering (S-MSC)—which applies mean shift clustering (MSC) to the short time Fourier transform (STFT) power spectrum for simultaneous de-noising and extraction of time-varying harmonic components—is proposed for the autonomous analysis of non-stationary vibration signals. A second pre-processing approach, termed Gaussian mixture model operating state decomposition (GMM-OSD)—which uses GMMs to cluster multi-modal vibration signals by their respective, unknown operating states—is proposed to address multi-modal non-stationarity. Applied in conjunction with S-MSC, these two approaches form a robust and unsupervised pre-processing framework tailored to the types of signals found in modern engineering systems. The final approach proposed in this thesis is a degradation detection and fault prediction framework, termed the Bayesian one class support vector machine (B-OCSVM), which tackles the key knowledge gaps pertaining to unsupervised parameter and degradation threshold estimation by re-framing the traditional fault detection and degradation modeling problem as a degradation detection and fault prediction problem.

Validation of the three aforementioned approaches is performed across a wide range of machinery vibration data sets and applications, including data obtained from two full-scale field pilots located at Toronto Pearson International Airport. The first of which is located on the gearbox of the LINK Automated People Mover (APM) train at Toronto Pearson International Airport; and, the second which is located on a subset of passenger boarding tunnel pre-conditioned air units (PCA) in Terminal 1 of Pearson airport. Results from validation found that the proposed pre-processing approaches and combined pre-processing framework provides a robust and computationally efficient and robust methodology for the analysis of non-stationary vibration signals in unsupervised CBM. Validation of the B-OCSVM framework showed that the proposed parameter estimation approaches enables the earlier detection of the degradation process compared to existing approaches, and the proposed degradation threshold provides a reasonable estimate of the fault manifestation point. Holistically, the approaches proposed in thesis provide a crucial step forward towards the effective implementation of unsupervised CBM in complex, modern engineering systems.

Acknowledgements

I would like to express my deepest gratitude and appreciation to my supervisor, Prof. Sriram Narasimhan. This journey simply would not have been possible without his technical expertise, patience and compassion.

I am grateful to Dr. Jinane Harmouche, Dr. Ali Ashasi-Sorkhabi, Prof. Guru Prakash and Prof. Jerome Antoni for providing their expertise, guidance and support throughout. In particular, I want to thank them all for helping me navigate through their respective areas of expertise. I'd also like to thank my committee members Prof. Wilson Wang, Prof. Kaan Erkorkmaz, Prof. Giovanni Cascante and Prof. Chris Bachmann, for dedicating their time to be a part of this process.

To Mike Riseborough and the Greater Toronto Airports Authority - thank you for the continued support and funding throughout the years. This longstanding partnership created many unique opportunities to apply this research in the real-world that would not have been possible otherwise. Similarly, I am grateful for the funding provided for this research through NSERC-CRD.

I also want to express my sincere gratitude to the all of the members of the Structural Dynamics, Identification, and Control (SDIC) lab for their friendship and support: Dirk, Kevin, Pampa, Marshal, Dylan, Rajdip, Roya, Nick, Evan, Steve and Alex.

Lastly, I would like to thank my family and friends for all of their love and support. Each of you has played an integral part in this process.

Dedication

To Mom and Dad—for everything

To Grandma

To Harold,

Medora,

& Cookie.

To Nina—if all this digging

had led me to nothing else,

it would have still been worth it.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

As the complexity of engineering systems continue to increase over time, so do the demands associated with the maintenance and operation of these systems. Such systems are better understood as a "system of systems" as opposed to a singular system. One example is the automated people mover (APM), which is comprised of multiple, interconnected control systems and mechanical assemblies that work in unison to meet the real-time needs of the application. Achieving the desired level of reliability in these systems requires the use of new, robust and dynamic maintenance strategies that are capable of keeping up the dynamic operation of these systems [100]. One such strategy is condition-based maintenance (CBM), which has quickly evolved to become a powerful alternative to traditional maintenance strategies such as preventative maintenance or run-to-failure maintenance [148]. CBM is a predictive maintenance strategy in which maintenance actions are performed based on the current health of the asset, which is inferred through a process known as condition monitoring (CM). Tying maintenance actions directly to the real-time health of the asset makes CBM well suited for the monitoring of highly dynamic systems. In addition, CBM is commonly deployed within an online framework to further facilitate the timely visualization of critical information.

The analytical portion of a CBM framework can be divided into two main branches: a diagnostic branch which includes fault detection, and a prognostic branch which deals with forecasting and degradation modeling [86]. The goal in diagnostics is to detect or classify a fault as soon as it appears to allow for corrective maintenance to be performed in a timely manner, while the goal of prognostics is to make the most accurate long-term predictions on the future health of the system using the data available. An effective CBM strategy is one that is capable of detecting faults early in their development stage, while simultaneously providing operators and planners with the necessary detail and foresight required

to schedule maintenance actions. A third component—pre-processing—is typically applied prior to both of these steps in order to maximize the quality of information represented in the data. In most real-world applications, pre-processing is necessary for obtaining accurate detection and prediction results. Hence, the overall effectiveness of a CBM framework is largely governed by the degree to which these three components perform collectively [152].

1.1 Motivation

Pertaining to CBM analytics as a whole, research and interest—particularly in methods utilizing machine learning (ML), signal processing and statistical methods—has seen significant growth over the last two decades [148, 104, 61]. Coupled with the advent of big data and increasing ubiquity of sensing in industry, the resulting exponential growth of CBM knowledge over this period has helped establish a strong foothold for CBM within industry. To date, the current body of literature in CBM is predominantly comprised of solutions and methods to tackle the supervised problem [104]—or in other words, the case where specific knowledge (i.e. labelled data, operational characteristics, failure behaviour) is assumed to be known. With respect to fault detection, the state-of-the-art in supervised ML approaches have already achieved near-perfect performance across most public validation data sets [209]. From a practical standpoint, however, these methods cannot be readily applied to many complex or novel engineering systems for which this pre-requisite information is not available.

Related issues also exist within the literature pertaining to prognostics and degradation modeling [61], where a central focus is to obtain an estimate of the remaining useful life (RUL) of a system [52]. In relation to maintenance and planning, an accurate RUL estimate provides operators with invaluable information that can be used towards optimizing maintenance actions and replacement scheduling. However, the crucial failure threshold used by most methods to obtain the RUL is almost always determined using prior information such as the failure histories of similar units or sensor measurements from a prior faulty state [102, 61]. The near-absolute requirement for historical data in current prognostic approaches inhibits the deployment of these methods in a wide range of real-world applications, especially those involving bespoke systems or systems where little to no prior knowledge is available.

In relation to pre-processing, the dynamic (i.e. non-stationary) operational profile of many modern systems requires the use of specialized approaches tailored to the analysis of the signals obtained from such systems. While many methods have been developed

and successfully applied to the analysis of such signals [138, 2, 15], these methods are typically computationally expensive, and require parameter tuning, domain expertise or prior knowledge in order to be effective. Hence, existing pre-processing methods for non-stationary signals are not well-suited either to the present research problem.

Achieving an effective realization of CBM on these types of novel systems requires the development of new approaches to address each of the aforementioned knowledge gaps within these research areas. The specific challenges of this research problem—in conjunction with the practical limitations of existing supervised approaches—form the motivation behind the development of computationally efficient, robust, and self-sufficient (i.e. unsupervised) methods for CBM.

1.2 Overarching Objectives

The overarching goal of this research is to develop novel methods to enable the effective realization of CBM within unsupervised, real-world settings. The aforementioned scientific and technical challenges of this application will require the development of robust statistical, machine learning and signal processing algorithms that can be applied to the pre-processing, fault detection and degradation modeling of non-stationary machinery vibration signals, without the need for prior knowledge of the system. In addition, to enable efficient and large-scale CBM deployment, these algorithms must be designed with practical considerations, such as computational complexity and compatibility with IoT in mind.

1.3 Overall Methodology

The overall methodology to achieve these objectives are summarized below, and contextualized within the unsupervised CBM framework in Fig. 1.1:

1. Development of an unsupervised and application-agnostic time-frequency approach for pre-processing and vibration analysis of non-stationary machinery vibration signals

2. Development of a robust, unsupervised degradation detection and fault prediction framework applicable to rotating machinery systems with no historical failures

3. Development of an unsupervised pre-processing approach for analysis of systems with multiple operating states

4. Validation of the proposed approaches using data collected from the industrial test beds located at Toronto Pearson International Airport as well as public machinery datasets

A comprehensive version of the flowchart in Fig. 1.1 and detailed description of the overall methodology is presented in section 4.2.1.

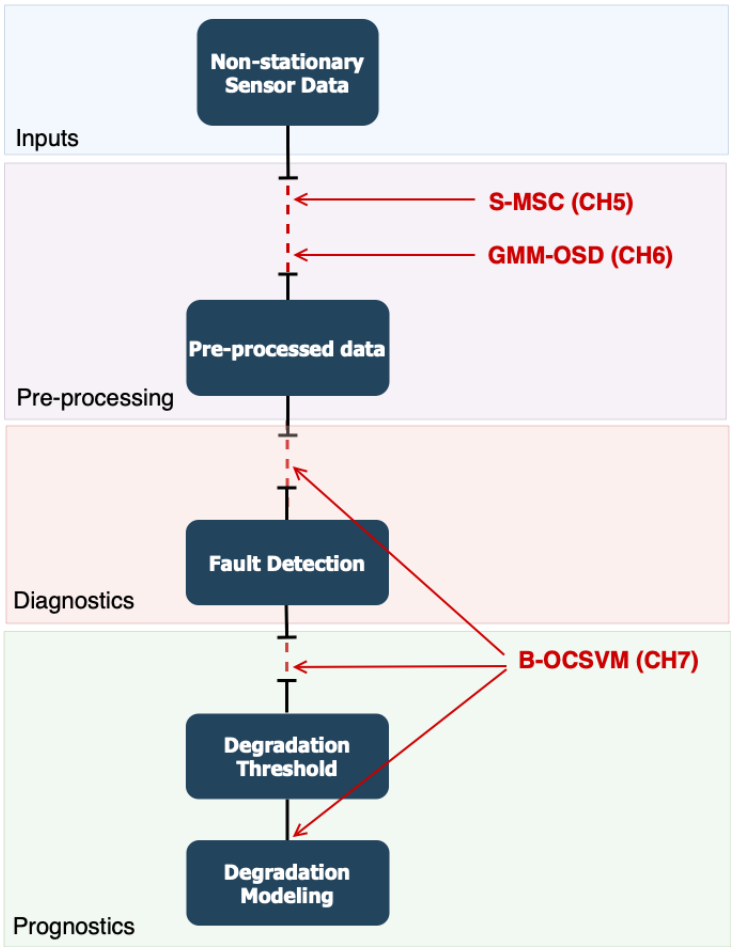

Figure 1.1: Contextualization of overall methodology with respect to unsupervised CBM; disconnects represent current gap areas, while red dashed lines and corresponding labels indicate the proposed solution(s) in this thesis to bridge each gap.

1.4 Organization of Thesis

This thesis is organized as follows:

- Chapter 1 presents a high-level overview of the current research problem, motivating statements, and a summary of the overall solution methodology.
- Chapter 2 presents the overarching theoretical background for the thesis, beginning with a description of the CBM architecture and it's various constituent components. The CBM architecture is used to contextualize the fundamental concepts presented afterwards within signal processing, machine learning, statistics and prognostics.
- Chapter 3 presents a review of literature on the state-of-the-art in CBM, specifically in relation to unsupervised CBM. The chapter begins by identifying the bodies of research within CBM relevant to this thesis, which is followed by a discussion of literature within three main areas: pre-processing and vibration analysis, fault detection and degradation modeling. The chapter concludes by summarizing the key knowledge gaps in the existing body of literature with respect to the effective realization of unsupervised CBM.
- Chapter 4 presents the proposed methodology to address the key knowledge gaps identified in Chapter 3. The main purpose of this chapter is to provide the reader with a clear understanding of the relationship between the different bodies of work within this thesis, as well as their relationship to the unsupervised CBM framework. The chapter begins with a detailed description of the specific research problems and presents a high level description of the proposed methodology to tackle each issue. The chapter concludes with a overview of the datasets used for validation in this thesis, including a description of the two field pilots located at Toronto Pearson International Airport.
- Chapter 5 presents the Spectral Mean Shift Clustering (S-MSC) time-frequency approach for non-parametric pre-processing and vibration analysis of non-stationary signals along with validation results obtained using the Pearson LINK APM gearbox, CMMNO'14 wind turbine [29] and Surveillance 8 Safran engine [8] datasets.
- Chapter 6 presents the Gaussian Mixture Model Operating State Decomposition (GMM-OSD) clustering approach for unsupervised pre-processing of vibration signals obtained from multi-operating state machinery along with validation results obtained using the Pearson LINK APM gearbox and Terminal 1 PCA unit datasets.

- Chapter 7 presents the Bayesian One Class Support Vector Machine (B-OCSVM) framework for unsupervised early degradation detection and fault prediction along with validation results obtained using the IMS [101], FEMTO [133] and C-MAPSS [163] run-to-failure datasets.
- Chapter 8 summarizes the main conclusions and key contributions of the thesis, followed by recommendations for future work.

Chapter 2

Background

At a holistic level, a fundamental understanding of the CBM framework can be achieved using the analogy of the modern day computer. In the simplest sense, the modern day computer can be thought of as a set of inter-connected sub-assemblies, software and instructions that work together to perform various functions in real-time, including interpretation, analysis and storage of data, visualization via graphics, along with providing connectivity to the internet or other external devices. The typical CBM framework functions in the same capacity - containing elements responsible for data collection, interpretation, analysis, visualization and connectivity that work together to meet the needs outlined by operators or specific policies. Consequently, the theoretical background for CBM has roots in many areas of engineering, including instrumentation, signal processing, machine learning, statistics, prognostics, Internet-of-Things and infrastructure management.

The scope of work presented in this thesis is focused on the analytical component of CBM, and specifically, on the development of new approaches and methods for the effective realization of unsupervised CBM. This chapter will provide the relevant, fundamental background for CBM and it's constituent components, beginning with an overview of the CBM architecture. Following the introduction of CBM, overarching concepts relevant to the analytical component of CBM from signal processing, machine learning, statistics, prognostics and machinery vibrations are presented. Specific concepts to each proposed CBM tool will instead be presented in the corresponding chapter.

2.1 Condition-based Maintenance

Condition-based maintenance, or CBM, is a predictive maintenance strategy in which maintenance actions are performed on an as-needed basis, as opposed to the scheduled intervals commonly found in traditional preventative maintenance approaches [86]. In CBM, maintenance actions for a particular asset are determined based on the current health, or condition of the asset. In practice, the current health of an asset is often inferred through some surrogate measure (i.e. sensor data) that can be obtained while the system is in operation. This process of active health tracking is known as condition monitoring (CM). CM is an integral component of any CBM framework, and the critical role it plays within the overall framework will be discussed in greater detail in the Section 2.2.

A high-level depiction of typical CBM architecture is depicted in Fig. 2.1, in which primary functions in a CBM framework are shown in sequential layers. The first level of a CBM framework is the input layer. In most data-driven CBM frameworks, the primary input is sensor data. Where applicable, historical records, such as any information pertaining to recorded failures, and SCADA data (i.e. operational status, speed, etc.) can also be provided as inputs. These inputs are typically passed through a crucial pre-processing step. The broad role of the pre-processing step is to maximize the readability of critical information in the input data before passing it along to the subsequent analytical component of the CBM framework. The analytical component of a typical CBM framework can be categorized into one of two main branches: a diagnostic branch which deals with the detection and classification of anomalies or faults; and, a prognostic branch which is concerned with forecasting and predictive modeling. While the primary functions of these two branches may differ at a fundamental level, strong synergies between these two components are often found at the core many successful CBM implementations. For example, outputs from the diagnostic branch (i.e. condition indicators, ML classification results) are commonly used as inputs to some prognostic models. Typical outputs of the analytical layer in a CBM framework include maintenance alarms, which are generated to alert the appropriate personnel when a fault is detected, and prognostic metrics such as remaining useful life and reliability. The final layer of the CBM framework is the decision support layer, which utilizes the outputs from the analytical layer to make informed maintenance and planning decisions. Alarm information can be used to trigger immediate maintenance actions or for maintenance scheduling, while predictions can be used for long term planning and scheduling [148].

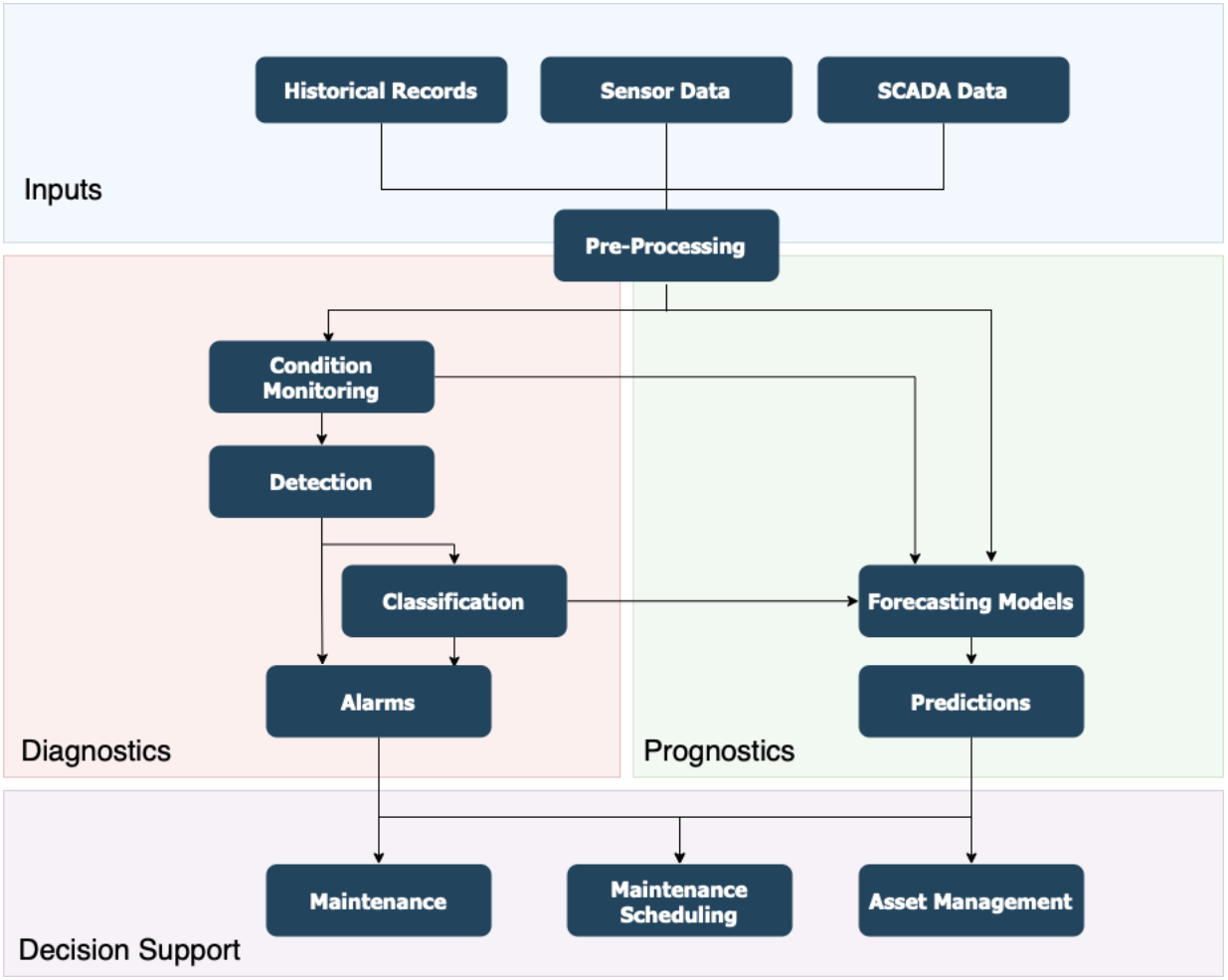

Figure 2.1: Typical CBM Architecture

2.1.1 Scope and Design of a CBM Framework

The extent of the diagnostic and prognostic capabilities in a given CBM framework are governed by a combination of user needs and practical constraints. Examples of user or application-specific needs include the type of asset to be monitored, the level of detection sensitivity required, or the long term goals of the end-user. User needs may impose requirements on the type and quality of sensors employed, or the depth of diagnostic and prognostic analysis required. The degree to which these specific user needs can be addressed are typically governed by practical constraints, such as implementation cost, installation requirements, and synergy with existing information technology (IT) and operational technology (OT) systems. Hence, there are often several trade-offs that need to be in the design process between ease of implementation and level of comprehensiveness: a generalized (i.e. "plug-and-play") framework will result in a low implementation cost, but will typically lack the sensitivity and comprehensiveness offered in a framework that has been fine tuned using expert knowledge. As new developments in CBM research continue to introduce new tools to address existing knowledge gaps, or improve upon the robustness of existing tools, this trade-off will diminish. It follows then, that the effectiveness of a CBM framework can be inferred by the degree to which it meets the specific needs of user or application, while staying within the imposed constraints.

Fig. 2.2 illustrates this CBM framework design process using the example application of a user who wants to deploy motor bearing monitoring across all of the units in their fleet. In Fig. 2.2, the user provides their specific needs, which include the determining the optimal replacement interval, detection of motor bearing faults, large-scale deployment of monitoring across the entire fleet, complete with the ability to visualize CBM information online and generate alarms automatically when a fault is detected. The CBM design begins by proposing the optimal solutions to meet each of these needs: vibration monitoring to maximize sensitivity to bearing faults, coupled with wireless sensing and remote connectivity via an IoT platform to provide fleet-wide coverage. Cost constraints impose a reduction on the program scope to limit monitoring to critical units only. Lack of historical failure data requires the analysis to be performed with methods which do not rely on such data (i.e. unsupervised methods). Finally, security constraints forces the design to adopt an LTE-enabled IoT implementation, and the subsequent increased costs of data transmission forces the use of edge computing and computationally efficient methods.

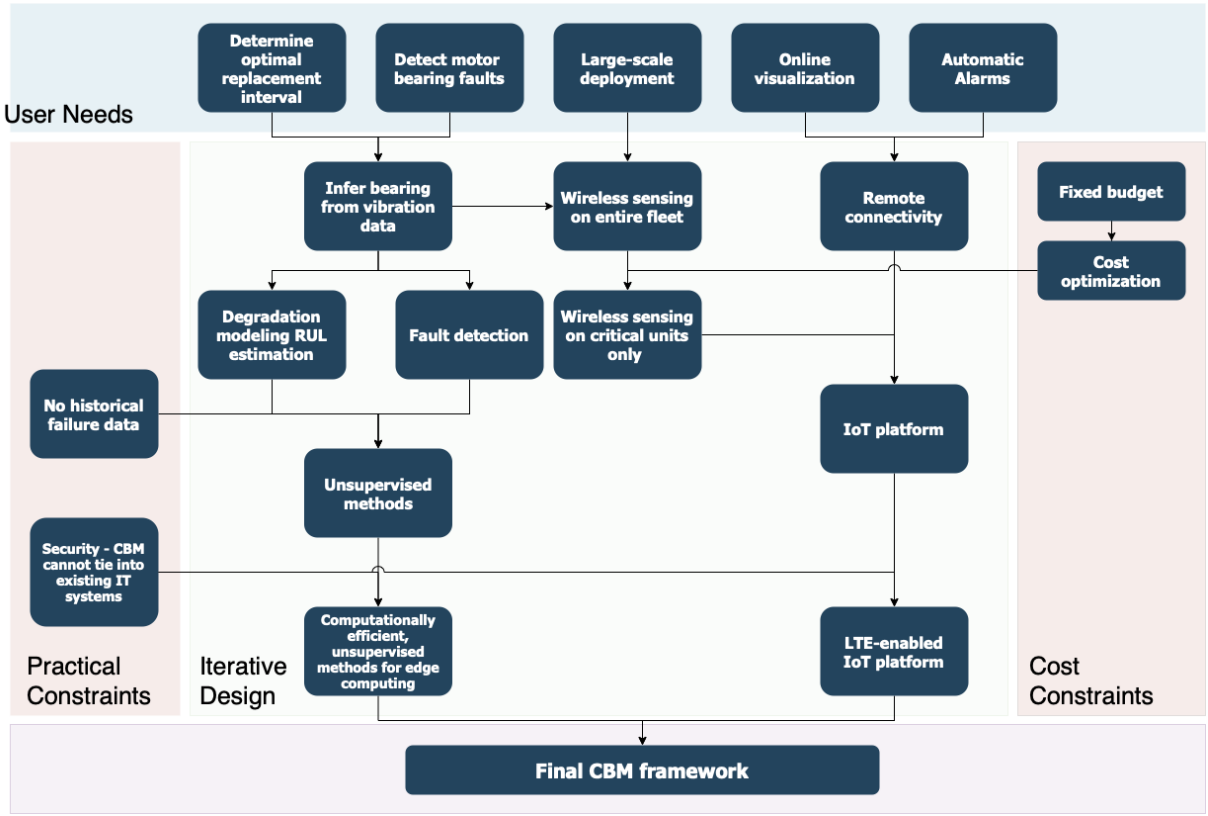

Figure 2.2: Example of CBM framework design process for fleet-wide motor bearing monitoring application

2.2 Condition Monitoring

In most data-driven CBM frameworks, CM is the primary method employed to track and infer the current health state of a monitored asset. For the majority civil and industrial assets, the current health of the asset cannot be inferred directly. In such cases, the most common approach is to infer the health of the asset using surrogate measures known as condition indicators (CIs), which are extracted from sensor data through the feature extraction process using signal processing, time series analysis or machine learning techniques [59, 184]. Prior to monitoring, a typical CM process will include a training phase in which extracted CIs are used to construct statistical or ML models which reflect the baseline (typically healthy) health of an asset. Once these baseline models are established, the monitoring phase of CM begins. During monitoring, CIs extracted from new observations of sensor data are classified as healthy or anomalous using the baseline models, in a process known as fault detection. These general steps of CM are shown in Fig. 2.3.

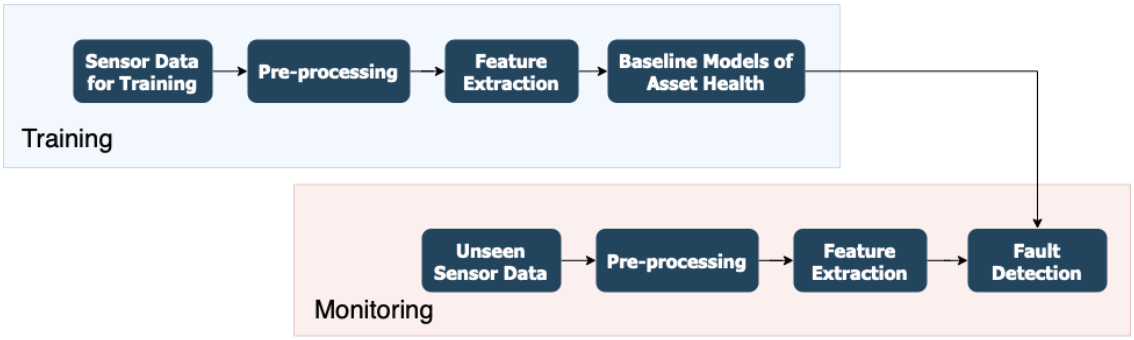

Figure 2.3: General steps of condition monitoring

2.2.1 Influence of CM on Overall CBM Framework

Beyond its role in monitoring and fault detection, CM has significant implications for the rest of the CBM framework. The type and nature of data pre-processing employed and choice of CIs has a significant impact on the subsequent diagnostic and prognostic analyses which use the outputs of CM as inputs.

2.2.2 Real-time Online CM

CM can be implemented at various levels of time localization and time resolution. In addition, the setting in which CM results are made accessible (i.e. offline vs. online) to the user can also vary framework-to-framework. The purpose of this subsection is to provide disambiguation and differentiation between two commonly misused terms in CM: "real-time" and "online". While these terms are often used interchangeably in CBM literature, there are several important distinctions to be made.

In the context of CM, the term "real-time" refers to when CM is employed in a fully continuous manner - sensor data is collected either continuously or at closely spaced intervals, and CI extraction and fault detection are performed immediately as new data becomes available. Consequently, the information obtained during real-time monitoring provides the highest level of time localization and time resolution. On the other hand, the term "online" is reference to the setting in which CM information is made available to the end-user. "Online" in this case means that results are accessible by the user over the internet or equivalent wireless connection. Typically, an online implementation will enable end-users to access real-time CM information with the minimal time delay. Due to this implication, the term "online" is often used interchangeably with "real-time" in literature. In this body of work, the phrase "real-time online monitoring" refers to the scenario in which CM is performed continuously, and the resulting information is made accessible to the end-user immediately.

2.3 Signal Processing

In practice, data captured by a sensor mounted on an asset is contaminated by a number of processes along the transmission path before it reaches the data acquisition device, including various sources of noise and other transmission path effects. In order to mitigate these negative effects, signal processing is typically applied to pre-process the raw sensor data prior to analysis in order to de-convolve the signal and extract the buried, useful information. Signal processing is a field of electrical engineering concerned with the manipulation of signals such as sound, images and measured signals [60]. The types of signals considered in this thesis consist exclusively of digitized sensor measurements. Hence, the theoretical background presented hereafter pertains to the sub-field of signal processing concerned with the analysis of digital signals known as digital signal processing (DSP). Pre-processing in CBM uses a wide range of signal processing tools, including filtering, time domain-based techniques, frequency-domain-based techniques and time-frequency-domain-

based techniques. The following sections will provide the general theoretical background for signal processing in the context of CBM.

2.3.1 Basic Signal Assumptions

In this thesis, the signals to be analyzed consist solely of sensor measurements, which obtain a discrete representation of an underlying time-continuous process by sampling said process repeatedly at specified time intervals. Let $x(t)$ represent a time-continuous process, where t is a continuous time variable. Then, the discrete representation of $x(t)$, as obtained through evenly-spaced sensor measurements can be written as $x[n]$. The relationship between $x(t)$ and $x[n]$ is shown in Fig. 2.4 and can be expressed mathematically as

$$x[n] = x(\frac{n}{f_s}) \qquad n = 1, 2, ..N \tag{2.1}$$

where, n is the quantized time variable or sample number, N is the total number of samples, and f_s is the sample rate. The sample rate f_s is inversely proportional to the time difference Δt between any two consecutive samples, given by

$$f_s = \frac{1}{\Delta t} \qquad \Delta t = t[n] - t[n-1] \tag{2.2}$$

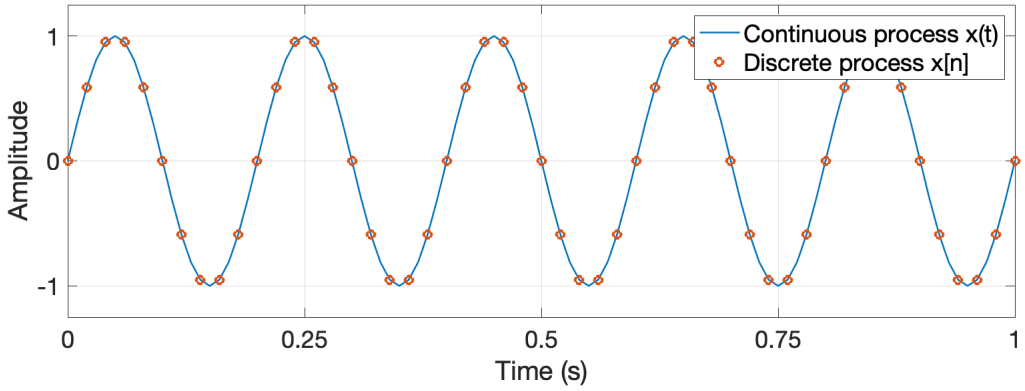

Figure 2.4: Continuous vs. discrete signal

In practice, it is common to assume that each measurement $x[n]$ consists of the signal of interest, denoted by $s[n]$, contaminated by some additive noise processes $\epsilon[n]$, resulting in

$$x[n] = s[n] + \epsilon[n] \tag{2.3}$$

In Eq. 2.3, each observation of $\epsilon[n]$ is assumed to be uncorrelated. In addition, the methods proposed in thesis will assume that the true probability distribution of ϵ is unknown. The following subsections will present several quintessential concepts from DSP.

2.3.2 Stationarity

In DSP, a stationary process is a statistically time-invariant process [170]. In other words, the statistical properties of a stationary process $x(t)$ would be the same at any given time t. Machinery vibration signals contain many examples of both stationary and non-stationary signals. To understand stationarity intuitively, consider the example of a machinery vibration signal collected from an APM gearbox shown in Fig. 2.5. The APM signal shown here contains many examples of stationarity and non-stationarity. Depending on the definition, the signal in Fig. 2.5 could be considered stationary or non-stationary. The signal in Fig. 2.5 can be considered stationary with respect to the mean. However, it cannot be considered stationary with respect to the variance. Stationarity is an integral concept in vibration analysis, as it often determines the type of tools that would be valid for analyzing a particular signal. Mathematically, a strict-sense stationary process satisfies the following relationship

$$p_{x(t)}(\alpha) = p_{x(t+k)}(\alpha) \qquad \forall t, k \in \mathbb{R} \tag{2.4}$$

where $p_{x(t)}$ is the probability density function of $x(t)$, α are the parameters of $p_{x(t)}$, and k is any number. In practice, the requirements for strict sense stationarity can rarely be satisfied by real signals. Hence, most DSP approaches will consider looser definitions of stationarity based on the mean, variance and co-variance.

A process can be classified with varying levels of stationary (i.e. first order stationary, second order or weak-sense stationary) [170]. The order is determined by the highest statistical moment of the process which satisfies the time-invariant property in Eq. 2.4. For example, in order for a process to be considered first order stationary, it must satisfy the following relationship

$$\mathbb{E}[x(t)] = \mathbb{E}[x(t+k)] \qquad \forall t, k \in \mathbb{R} \tag{2.5}$$

In other words, a first order stationary process is one in which the mean is constant. Similarly, it follows that a second order or weak-sense stationary process is one in which the joint probability density function of $x(t)$ and $x(t+k)$ is the same for all k such that

$$\begin{aligned} \mathbb{E}[x(t)] &= \mathbb{E}[x(t+k)] \qquad \forall t, k \in \mathbb{R} \\ \mathbb{E}[x(t)x(t+k)] &= \mathbb{E}[x(t+\tau)x(t+k+\tau)] \qquad \forall t, k, \tau \in \mathbb{R} \end{aligned} \tag{2.6}$$

The final type of stationarity that is required for vibration analysis is cyclostationarity. As the name implies, cyclostationarity refers to a signal that exhibits stationarity at periodic intervals [170]. Similar to the definition in Eq. 2.6, a process is considered cyclostationarity if it exhibits the same mean and joint probability density function at periodic intervals.

In general, the performance of methods which have been designed for analysis of stationary signals does not translate well to non-stationary signals. Analysis of non-stationary signals requires specialized tools that are able of adapting to or characterizing the time-varying properties of non-stationary signals. This class of tools forms the basis for much of the body of work contained in this thesis.

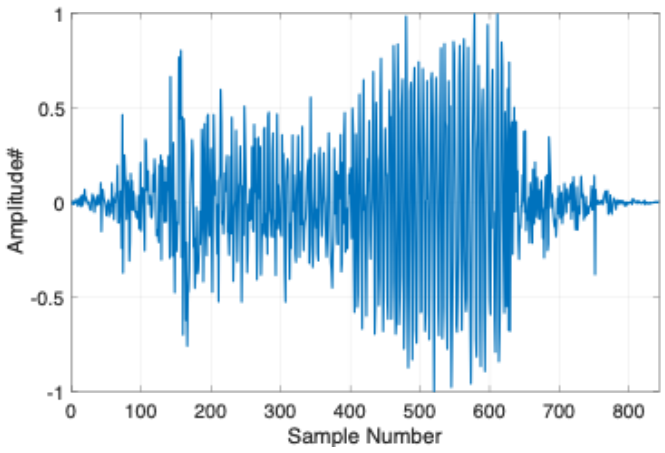

Figure 2.5: Example of non-stationary machinery vibration signal obtained from APM gearbox

2.4 Fourier Analysis

In CBM, the frequency content of a signal is commonly used for many applications including fault detection, filter design and system identification. Fourier analysis, which assumes that a signal can be expressed a sum of trigonometric or exponential functions of different frequencies, forms the basis from which most common frequency analysis methods are predicated upon [60]. Let $g(t)$ represent a periodic signal with period T such that

$$g(t) = g(t + mT) \qquad m \in \mathbb{Z} \tag{2.7}$$

Fourier Analysis states that $g(t)$ can be expressed as an equivalent Fourier series with the form

$$g(t) = \sum_{k=-\infty}^{\infty} C_k e^{jk\omega_0 t} \tag{2.8}$$

where $\omega_0 = 1/T$ is the fundamental angular frequency, and C_k is the Fourier series coefficient given by,

$$C_k = \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-jk\omega_0 t} dt \tag{2.9}$$

In 2.8, C_k represents the magnitude of each frequency component $k\omega_0$. Hence, Fourier analysis allows us to view the equivalent frequency representation of a periodic time signal. In practice, signals obtained from sensor measurements are rarely perfectly periodic. The Fourier series approximation in Eq. 2.8 can be extended for non-periodic signals as well by allowing $T \to \infty$. Let $f(t)$ represent a non-periodic signal with relationship to $g(t)$ given by

$$f(t) = \lim_{T \to \infty} g(t) \tag{2.10}$$

Bringing C_k to the LHS of 2.9 yields,

$$TC_k = \int_{-T/2}^{T/2} g(t) e^{-jk\omega_0 t} dt \tag{2.11}$$

Setting $k\omega_0 = \omega$ at $T \to \infty$ results in

$$\omega_0 = \frac{2\pi}{t}|_{T \to \infty} \to 0 \tag{2.12}$$

Eq. 2.12 transforms the discrete Fourier spectrum into continuous, replacing the summation by integral which results in $f(t) \to g(t)$. Then, the Fourier transform of the non-periodic signal $f(t)$ is obtained as $T \to \infty$ in Eq. 2.11

$$\begin{aligned} TC_k &= \lim_{T \to \infty} \int_{-T/2}^{T/2} g(t) e^{-j\omega t} dt \\ TC_k &= \int_{-\infty}^{\infty} [\lim_{T \to \infty} g(t)] e^{-j\omega t} dt \\ F(\omega) &= \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \end{aligned} \tag{2.13}$$

where $F(\omega)$, known as the spectral density, represents the frequency spectrum of $f(t)$. The reverse transformation, which can be used to re-construct the original time signal is referred to as the Inverse Fourier Transform (IFT), and is defined as

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \tag{2.14}$$

Oftentimes, the Fourier transform pair defined by Eq. 2.13 and 2.14 are expressed using the following notations

$$F(\omega) = F[f(t)] \qquad \text{or} \qquad f(t) = F^{-1}[F(\omega)] \tag{2.15}$$

2.4.1 Discrete Fourier Transform

The formulation used to arrive at Eq. 2.13 and 2.14 contain two assumptions that cannot be realized in practice. Namely, the signals obtained through sensor measurements are neither continuous functions nor infinite in length. Rather, the signals of interest are discrete and finite in nature. Applying the transforms in Eq. 2.13 and 2.14 to these signals requires yet another extension of the Fourier Transform known as the Discrete Fourier Transform (DFT) [60]. In the DFT, the infinite integrals of the FT are replaced by finite sums, and is expressed as

$$F(k) = \frac{1}{N} \sum_{n=0}^{N-1} f(n) e^{-j2\pi kn/N} \qquad k = 0, 1, ..., K \tag{2.16}$$

where k indicates the k^{th} discrete spectral component, N is the length of the input signal, $K = N/2$ is the total number of spectral components according to Nyquist theorem. Division by N ensures that the Fourier series components are scaled accordingly. The corresponding Discrete Inverse Fourier Transform is given by

$$f(n) = \sum_{k=0}^{N-1} F(k) e^{j2\pi kn/N} \tag{2.17}$$

In matrix notation, the DFT in Eq. 2.16 can be represented by the following matrix multiplication

$$\mathbf{F}_k = \frac{1}{N} \mathbf{W}_{kn} \mathbf{f}_n \tag{2.18}$$

where $\mathbf{F}_k$ is a vector of N frequency components, $\mathbf{f}_n$ is a vector containing the original input signal, and $\mathbf{W}_{kn}$ is a square matrix of unit vectors $e^{-j2\pi kn/N}$ whose angular orientation is a function of both k and n.

2.4.2 Fast Fourier Transform

In many modern computing packages, the DFT is often evaluated using the Fast Fourier Transform (FFT) algorithm [170]. The FFT is computationally efficient method that evaluates Eq. 2.16 and 2.17 under the principle that N is a power of 2. The FFT algorithm factorizes $\mathbf{W}_{kn}$ in Eq. 2.18 into $log_2 N$ matrices each requiring only N complex operations for multiplication, in contrast to the N^2 operations required for direct multiplication using $\mathbf{W}_{kn}$. Hence, the FFT algorithm reduces the overall computational complexity from N^2 to $Nlog_2 N$.

2.4.3 Nyquist Frequency, Sampling Rate and Frequency Resolution

Three crucial parameters in Fourier analysis are the Nyquist Frequency, the previously defined sampling rate f_s and the frequency resolution. The choice in each of these parameters will have a large impact on the type and quality of frequency information available for analysis.

Nyquist Frequency and Signal Bandwidth

According to the Nyquist theorem, the Nyquist frequency represents the highest frequency that can unambiguously represented in a sampled signal [170]. Frequency components above the Nyquist frequency will become indistinguishable due to aliasing, which refers to the phenomenon when a signal component becomes an alias or copy of another. The Nyquist frequency is equal to half of the sampling rate f_s. In practice, f_s is typically determined from a priori knowledge of the highest frequency component of interest in a given application. Fig. 2.6 illustrates the aliasing phenomenon using a simple 60 Hz cosine wave. According to Nyquist theory, a minimum sampling rate of 120 Hz would be required to capture the cosine wave without introducing aliasing effects. Fig. 2.6a illustrates the cosine wave sampled at 70 Hz. Fig 2.6b shows that due to aliasing, the waveform of the sampled points appears instead as a 10 Hz cosine wave.

In Fourier analysis, the signal bandwidth simply refers to the maximum range between the highest and lowest umambiguous frequency components in a signal, which for a raw unprocessed signal, is the range between zero and the Nyquist frequency. The term "bandwidth" is also used in filtering where it carries a slightly different meaning. To disambiguate between the two use cases, "bandwidth", when used with respect to Fourier Analysis will be referred to instead as "signal bandwidth".

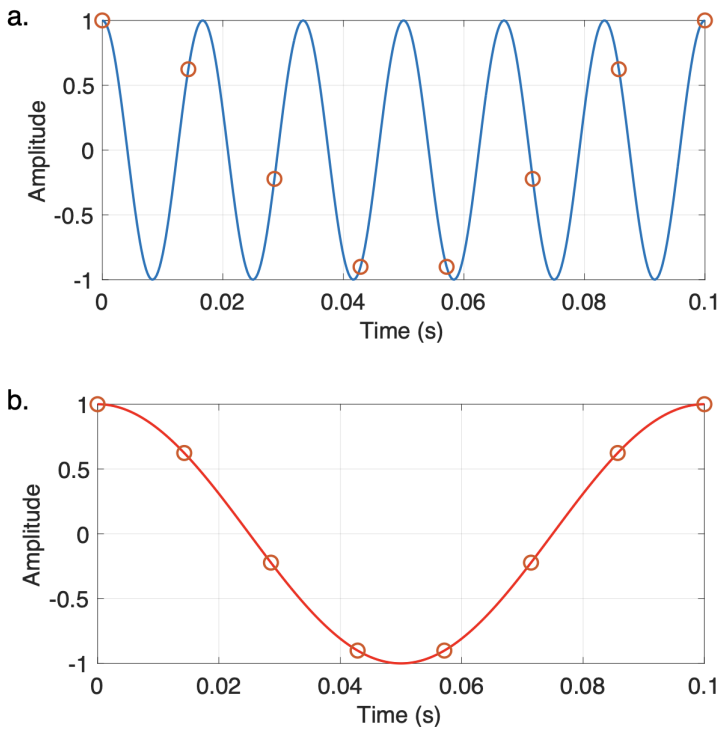

Figure 2.6: Example of aliasing. a) 60 Hz cosine wave sampled at fs = 70 Hz b) Corresponding waveform of sampled points aliasing as a 10 Hz cosine wave

Sampling Rate and Frequency Resolution

In addition to its relationship with the Nyquist frequency, the choice of f_s also controls the maximum frequency resolution available for analysis. Frequency resolution refers to the size of the interval between two consecutive frequencies in the Fourier spectrum, which is determined by the relationship f_s/N, where N is the length of the input signal to the DFT [170]. Frequency resolution represents the minimum required spacing for two frequencies to be distinguished uniquely in the DFT. Frequencies spaced closer together than the frequency resolution will appear as a single frequency in the DFT. Increasing the frequency resolution will allow for better differentiation of closely spaced frequencies, and is particularly important in many CBM fault diagnosis approaches which rely on the identification of fault frequencies that can manifest around harmonic components as closely spaced "sideband" frequencies. Hence, where practically feasible, oversampling of

the signal (i.e. sampling at a rate higher than two times the Nyquist frequency) is desirable in order to obtain the best possible frequency resolution for analysis.

2.4.4 Windowing

Under ideal conditions, the time signals analyzed using Fourier analysis will contain a perfect integer multiple of the period. In practice, this condition is rarely satisfied, and the start and end points of a sampled signal will contain discontinuities which will manifest as high frequency components above the Nyquist frequency in Fourier analysis. These artificial high frequency components will lead to aliasing and spectral leakage in the frequency spectrum. Spectral leakage, refers to the phenomenon in which the energy in one frequency is smeared into other frequencies, which makes it difficult to distinguish closely spaced frequencies [170]. Controlling or managing spectral leakage is particularly important when analyzing machinery vibration signals, which are often contain many closely spaced frequency components. The first purpose of windowing in DSP is to eliminate these discontinuities in the time domain of the signal in order to control spectral leakage. The second common application of windowing is for time-frequency analysis, which will be discussed in more detail in section 2.5.

Typically, a window is a function that is real-valued, smooth and symmetric over a chosen interval and zero otherwise. Hence, when a window function is applied to a signal, the symmetry property will force the start and end points of the signal within the windowed interval to be continuous. The two main parameters that need to be determined when designing a window function $w(n)$ are the length of the window L_w, and the type of window. The two main types of window functions used in analysis of machinery vibration signals are the Hanning and Dirichlet (rectangular) windows.

- **Hanning:** the Hanning window is a tapered cosine window that is commonly used to control spectral leakage, especially in the analysis of machinery vibration signals [170]. The Hanning window symmetrically tapers to zero at each end, which helps to minimize spectral leakage. The Hanning window can be expressed by the following equation

$$w(n) = 0.5(1 - cos(2\pi \frac{n}{L_w - 1}) \qquad 0 \leq n \leq L_w - 1 \tag{2.19}$$

- **Rectangular:** the rectangular window is the simplest type of window function, in which all points within the window are weighted equally (i.e. $w(n) = 1$). Unlike the Hanning window, the rectangular window does not help to control spectral leakage. Rectangular windows are commonly used in time-frequency analysis to capture

transients or short-lived events within a signal, which are difficult to capture with Hanning windows due to their smoothing effects.

Fig. 2.7 illustrates the time and frequency representations for the Hanning and Rectangular windows. The optimal choice of window type and length will vary depending on the type of signal and analysis required.

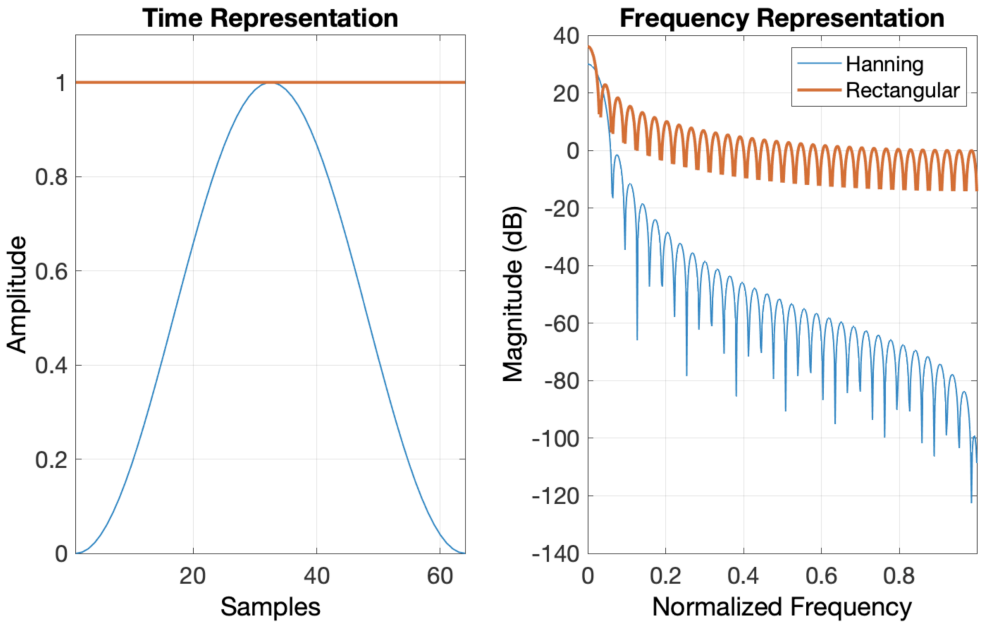

Figure 2.7: Time and frequency representations for the Hanning and Rectangular Window Functions

2.5 Time-frequency Analysis

Time-frequency (TF) analysis is an extension of Fourier analysis that focuses on the study of a signal in both the time domain and frequency domain simultaneously. TF methods are essential for analyzing non-stationary signals, such as those commonly obtained from rotating machinery. Contrary to Fourier analysis, which produces a 1-D output, time-frequency analysis produces 2-D outputs which are referred to as time-frequency representations. TF analysis can also be used to capture transients and other short-lived events within a signal

that would otherwise be obscured in regular Fourier analysis. The following sections will present several of the key TF methods which will be used throughout this thesis.

2.5.1 Short-time Fourier Transform and Spectrogram

The Short-time Fourier Transform (STFT) is the natural extension of the FT to the time-frequency domain [60]. In essence, the STFT is obtained by combining the FT with windowing. Thus, instead of calculating a single FT to represent the frequency content of the entire signal, the STFT calculates the FT of the signal at each specified window, allowing any short-lived phenomena or other non-stationary behaviour within the signal to be captured through some subset of windows. Mathematically, the discrete STFT $S(m,k)$ for a signal of length L is expressed as

$$\begin{aligned} S(m,k) = \sum_{n=0}^{N-1} x(n+mH)w(n)e^{-j2\pi kn/N} \\ M = \frac{L-N}{H} \\ m = 1,2,...,M \\ k = 1,2,...,K \end{aligned} \tag{2.20}$$

where m dictates the window number, k represents the k^{th} spectral component, $w(n)$ is a window function of length N, $K = N/2$ is the total number of spectral components as dictated by Nyquist theorem, H represents the shift length between two consecutive windows and M is the total number of windows. The most common representation of the STFT is the spectrogram, which is simply the squared amplitude of the STFT in Eq. 2.20, commonly expressed as

$$\hat{S}(m,k) \equiv |S(m,k)|^2 \tag{2.21}$$

2.5.2 Hilbert Transform and Envelope Analysis

Although inherently not a time-frequency method, the Hilbert Transform (HT) is widely used in tandem with filtering for time-frequency analysis of machinery vibration signals [153]. In vibration analysis, the primary function of the HT is to obtain the analytic signal, which can detailed information regarding the structure of the signal, such as the

instantaneous amplitude (i.e. envelope) or phase. This information is often useful for diagnosing machinery faults, which can manifest more clearly in within the envelope of a signal. Mathematically, the HT of a signal $x(t)$ can be simply expressed as the convolution of $x(t)$ with kernel $1/\pi t$

$$\mathcal{H}[(x(t)] = \frac{1}{\pi t} \int_{-\infty}^{\infty} \frac{x(t-\tau)}{\tau} d\tau \tag{2.22}$$

Using the definition of HT in Eq. 2.22, the analytic signal $\bar{x}(t)$ of $x(t)$ can be obtained by adding the imaginary component of the HT $\mathcal{H}x(t)$ to $x(t)$

$$\bar{x}(t) = x(t) + j\mathcal{H}(x(t)) \tag{2.23}$$

The instantaneous amplitude or envelope signal $|\bar{x}(t)|$ can then be obtained by simply taking the absolute value of Eq. 2.23. Fig. 2.8 below illustrates the envelope signal extracted from a convolved cosine signal.

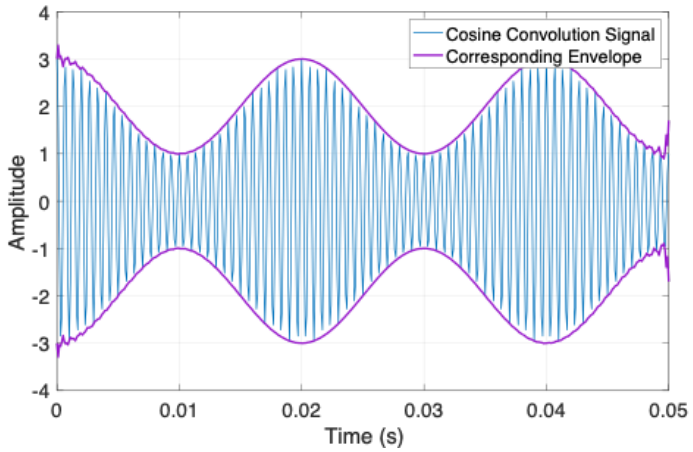

Figure 2.8: Example of envelope signal extracted using the Hilbert transform

2.5.3 Spectral Kurtosis

The spectral kurtosis (SK) is a statistical parameter that can be used for transient detection in the frequency domain [6]. Conceptually analogous to the kurtosis of a signal in the time domain, the SK is defined as the normalized fourth-order moment of the real part of

previously defined discrete short-time Fourier transform $|S(m,k)|$, given by:

$$SK(k) = \frac{\langle |S(m,k)|^4 \rangle}{\langle |S(m,k)|^2 \rangle} \tag{2.24}$$

where $\langle \cdot \rangle$ represents the time-averaging operator. Similar to the behaviour of kurtosis in the time domain, the value of SK is nominal at frequencies containing only stationary Gaussian components, and high for frequencies containing transient, or non-Gaussian behaviour. The ability of SK to perform transient detection at the per-frequency level makes it well suited to the detection of various narrow-band machinery fault signatures [6]. Fig. 2.9 below illustrates this use of SK for the detection of a narrow band linear chirp buried in dense, Gaussian noise. A popular extension of SK is the Kurtogram, which essentially computes the SK as a function of the window length in addition to frequency and provides a convenient tool for parameter selection in time-frequency analysis [7].

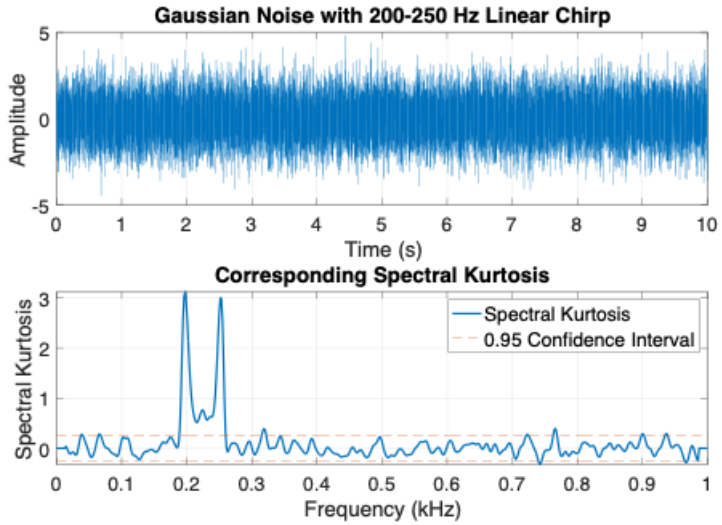

Figure 2.9: Example using spectral kurtosis for detection of a narrow band linear chirp buried in Gaussian noise

2.5.4 Time Synchronous Averaging

Time synchronous averaging (TSA) is a pre-processing technique used for the isolation of periodic signal components in noisy or non-stationary data, and is widely used in the

diagnosis of machinery faults [15]. The principle of TSA is to resample a signal in the time domain into the angular domain with respect to the phase of a particular component of interest (i.e. as a function of shaft rotation). In doing so, any signal components not synchronized with the phase of the component of interest will be filtered out of the signal. TSA requires angular reference signal that is synchronized with the time signal (i.e. phase-locked). In practice, this angular reference is typically obtained using tachometer measurements; however, tachometer-less (albeit computationally expensive) estimation approaches also exist within the literature [8].

2.6 Filtering

In the broad sense, the main purposes of filtering are signal separation and signal restoration [170]. Signal separation, which is the primary type of filtering employed in this thesis, refers to the process of isolating or separating frequency components of interest (i.e. the original signal) from noise and sources of interference encountered along the transmission path. Signal restoration refers to the process of recovering or reconstructing a signal that has been distorted i.e. de-blurring an image that is out of focus. In essence, a filter is a process which takes an input signal and produces an output that is either a linear or non-linear combination of the input. Additionally, filters mostly commonly operate within the time domain - that is to say that the input and output signals of a filter are in the time domain. This thesis will employ both linear and non-linear filtering methods that act within the time, frequency and time-frequency domains. The filter theory presented hereafter will focus on digital filtering practices for signal separation in digitally sampled signals.

2.6.1 Linear Filters

A linear filter can be understood as a filter in which the output signal is a linear combination of the input signal [170]. Let $\mathbf{x}$ represent an arbitrary input signal of length N. Then, the output $y[n]$ obtained by passing $\mathbf{x}$ through a filter $\mathcal{L}$ is expressed by the following relationship

$$y[n] = \mathcal{L}_n\{\mathbf{x}\} \tag{2.25}$$

In Eq. 2.25, the filter $\mathcal{L}_n\{\}$ is expressed as a real-valued function with value at every sample n. A filter $\mathcal{L}$ is linear if satisfies the superposition principle for linear systems given

by

$$\begin{aligned} \mathcal{L}_n\{a\mathbf{x}\} &= a\mathcal{L}_n\{\mathbf{x}\} \\ \mathcal{L}_n\{\mathbf{x}_1 + \mathbf{x}_2\} &= \mathcal{L}_n\{\mathbf{x}_1\} + \mathcal{L}_n\{\mathbf{x}_2\} \end{aligned} \tag{2.26}$$

where a is a scalar. A linear filter can be described by it's impulse response function (IRF), step response or frequency response function (FRF) [170]. As the name implies, the impulse response of a filter is simply the output the filter produces when an impulse is provided as input. The frequency response can be obtained by taking the FT of the impulse response, and it follows that the step response can be obtained by integrating the impulse response. Each of these descriptions can be used to implement and understand the effect a given filter will have on an input signal. Fig. 2.10 illustrates the step, impulse and frequency response functions for a simple digital linear filter. These functions will be useful when defining the various types of filters hereafter. Mathematically, the IRF of a

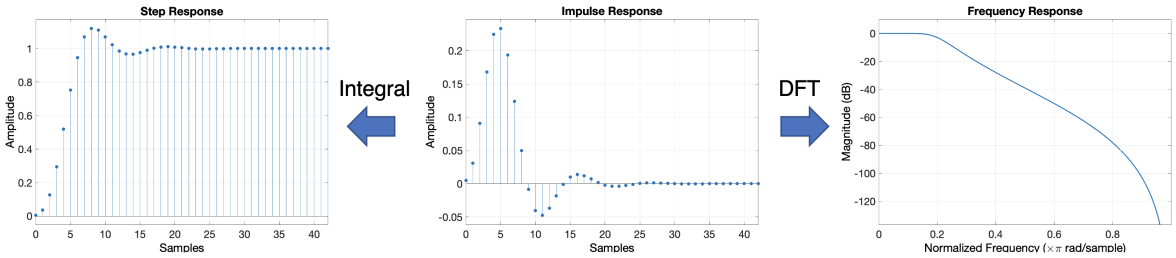

Figure 2.10: Example of a step, impulse and frequency response functions for a digital linear filter

filter is expressed as

$$\begin{aligned} h[n] &= \mathcal{L}_n\{\delta_n(\mathbf{x})\} \\ \text{where} \qquad \delta_n(\mathbf{x}) &= \begin{cases} 1 \text{ if } n = 0 \\ 0 \text{ if } n \neq 0 \end{cases} \end{aligned} \tag{2.27}$$

One method to implement a linear filter is to convolve the input signal with the IRF of a filter. In other words, the output signal of a filter can be understood as the weighted sum of the input signal. Filters implemented using convolution are also known as finite impulse response (FIR) filters. One other important type of implementation approach that builds upon the convolution in a more efficient manner is known as recursion. Recursion is the method employed by many common filters, including Butterworth filters. In recursion, the

next output $y[n]$ in a sequence is defined as a weighted combination of the previous inputs as well as the previous outputs. Mathematically, recursion can be expressed as

$$y[n] = \sum_{l=1}^{N} a_l y(n-1) + \sum_{k=0}^{M} b_k x(n-k) \tag{2.28}$$

where a_l and b_k are the called the recursion coefficients. Recursive filters are defined by their recursion coefficients, as well as the filter order (typically $N = M$). Filters implemented using recursion are also known as infinite impulse response (IIR) filters. IIR filters are generally more efficient than their FIR counterparts, but this comes at the cost of being more complex to design.

2.6.2 Common Types of Frequency Response Functions

In engineering, filters are often described by their FRF. The FRF provides an intuitive way to understand which frequencies a given filter will reject or admit, as well as the level of attenuation applied to each frequency band. The band of frequencies admitted by a filter is known as the passband, while the band of frequencies rejected by a filter is known as the stopband. The boundary(s) between these two regions is known as the cutoff frequency f_c.

The four most common types of filters as classified by their FRFs are low-pass, high-pass, band-pass and stopband filters. These four types of filters are shown in Fig. 2.11. Beyond rejecting the frequencies in the stopband, most filters will also produce some form of attenuation or ripple effect in the passband as well, which can be seen in all four FRFs in Fig. 2.11.

- **Low-pass filters:** low-pass filters (Fig. 2.11a) admit frequencies below f_c and reject frequencies above f_c
- **High-pass filters:** high-pass filters (Fig. 2.11b) admit frequencies above f_c and reject frequencies below f_c
- **Bandpass filters:** bandpass filters (Fig. 2.11c) admit frequencies in the band bounded by the cutoff frequencies f_{c1} and f_{c2}, and rejects all others
- **Bandstop filters:** bandstop filters (Fig. 2.11d) admits all frequencies except for those in the band bounded by the cutoff frequencies f_{c1} and f_{c2}.

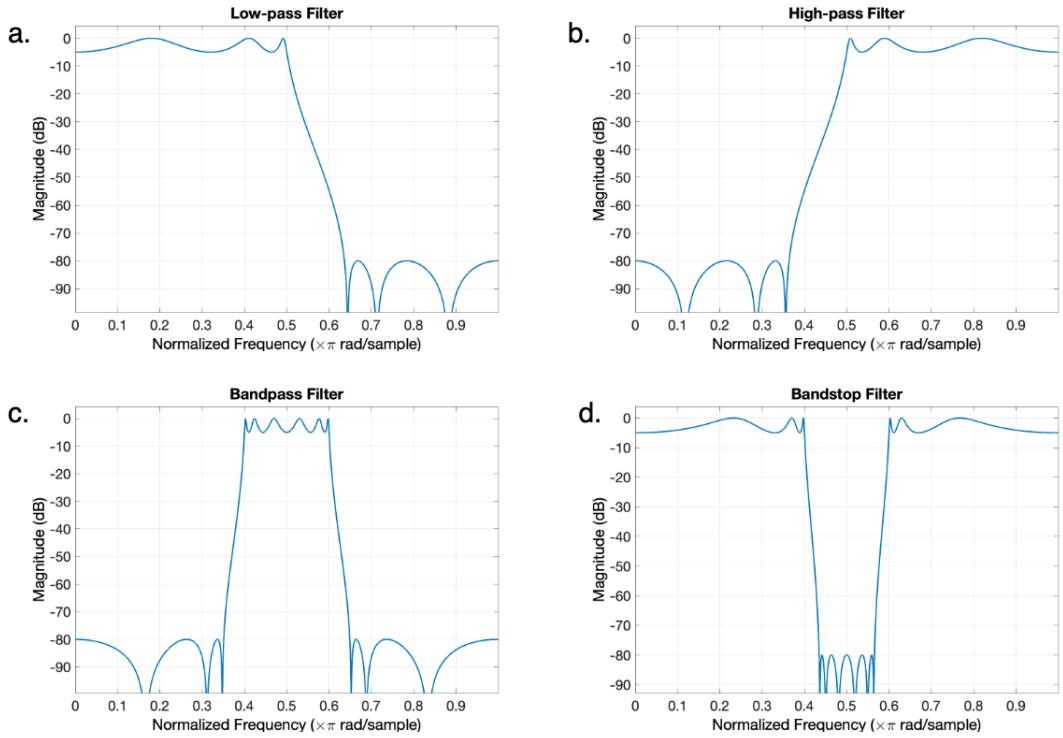

Figure 2.11: Common FRFs a) Low-pass filter b) High-pass filter c) Bandpass filter d) Bandstop filter

2.6.3 Butterworth Filters

One of the most widely used recursive linear filters is the Butterworth Filter [20]. Butterworth filters have a unique property in their frequency response in which the passband is maximally flat, which comes at the expense of a shallower slope in the transition region between the passband and stopband compared to other filters (i.e. elliptic) [170]. Hence, Butterworth filters are frequently used in engineering applications due to their ease of design and intuitiveness. The recursive form of the Butterworth filter is identical to the expression derived in Eq. 2.28, and hence, Butterworth filters are defined by the set recursive coefficients b_l and a_k and the filter order (commonly $N = M$). Filter order controls the steepness of the roll-off at the cutoff frequency. The effect of filter order on a low-pass Butterworth filter is shown in Fig. 2.12. Increasing the filter order will result in a steeper roll-off. However, increasing the filter order past a certain point will begin to introduce

ripples in the passband, which would negate the main advantage of the Butterworth filter. The method to solve for the Butterworth filter coefficients can be found in **Appendix A**.

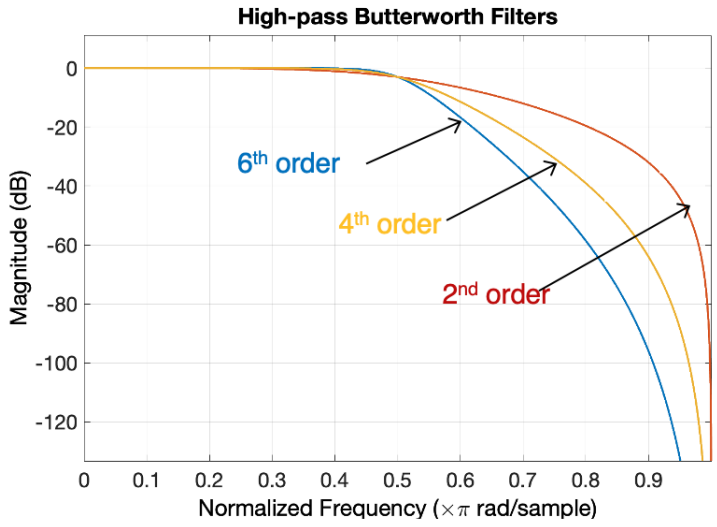

Figure 2.12: Effect of filter order on low-pass Butterworth filter

2.6.4 Non-Linear Filters

As the name implies, a non-linear filter can be used to describe any filter whose output is not a linear function of its input. In other words, a non-linear filter does not satisfy the superposition principle of a linear system defined by Eq. 2.26. Unlike linear filters, the frequency response of a non-linear filter cannot be easily characterized. A common application of non-linear filters is for de-noising [170]. Specifically, non-linear filters can be used to remove non-additive noise processes from a signal. An intuitive example of one such filter is the median filter, which is commonly used for signal smoothing. Fig.2.13 illustrates how a median filter can be used to recover a sine wave contaminated by multiplicative noise. Another example of a pseudo-non-liner filter introduced earlier in this chapter is TSA [15].

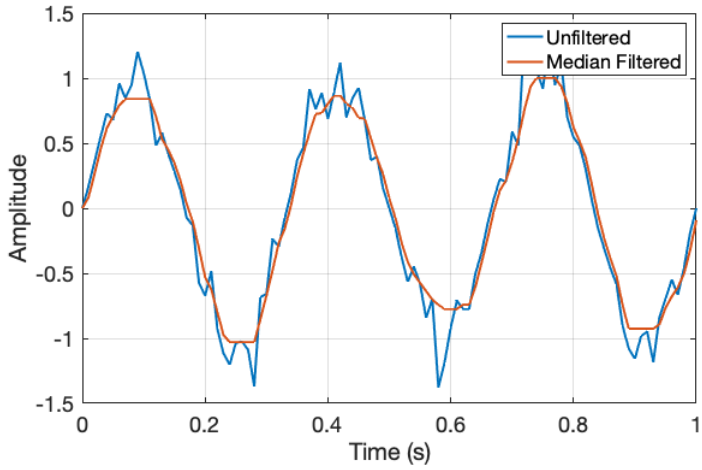

Figure 2.13: Removal of multiplicative noise using median filter

2.7 Machine Learning and Statistical Analysis

Machine learning (ML) and statistical analysis tools have been widely implemented in all aspects of CBM and vibration analysis, including feature extraction, fault detection, fault classification and degradation modeling [160, 17]. The field of statistics is often considered as the pre-cursor to modern ML, and as result, many methods from both fields today have dual classifications. Despite the vast sea of ML and statistical tools available today, the optimal choice(s) of tools will depend primarily on the application and corresponding constraints. Hence, achieving the optimal design or performance with these tools in a particular application requires both expert knowledge and careful consideration of application-specific parameters, needs and constraints. This section presents the overarching concepts used throughout this thesis within ML and statistical analysis. Concepts which are specific to a particular method will be presented in their corresponding chapters.

2.7.1 Classification of ML Methods

At the highest level, ML methods can be classified by the type of learning approach used. The learning approach of a method will dictate the type and level of information required by that method for training. The three types of learning approaches are: supervised, semi-supervised and unsupervised learning. The descriptions provided here are intended to provide a high-level understanding of the key characteristics and differentiating factors

between each classification.

Supervised Learning

Supervised learning assumes that data used for training a classifier is fully labelled, or in other words, that the case each point in the training set belongs to (i.e. healthy, faulty) is known a-priori. These labels are passed onto the classifier along with the data, and the goal of a classifier in supervised learning is to learn the best relationships between the data and it's assigned class label, and apply these relationships to the classification of unseen data. Hence, the common task in supervised learning techniques is classification i.e. determining the correct class assignment for a new observation [17]. Widely used examples of supervised ML tools including neural networks (NNs) [199] and support vector machines (SVMs) [194].

Unsupervised Learning

Contrary to supervised learning, unsupervised learning assumes the opposite scenario where the training data provided to the classifier contains no label information. Hence, the number and nature of the underlying classes in the training data are not known a-priori. Since unsupervised classifiers are unable to make use of labels for learning class relationships, these approaches typically make use of inducted biases (i.e. assume only one of class present in the training data) in order to develop decision rules. An explicit example of this is OC-SVM [166]. Another common class of methods within unsupervised learning is clustering, which is used to describe the process of organizing or grouping points in a data set such that the similarity of points to each other within a given group (i.e. cluster) is than to those in other groups [17]. An example of a common unsupervised clustering technique is K-means clustering [17].

Semi-supervised Learning

Semi-supervised learning bridges the gap between supervised and unsupervised learning. Semi-supervised learning involves the use of a partially labelled training set. For example, within a training set, complete labels may be available for only a subset of data points, which will require a semi-supervised classifier to use those labels to make inferences on the remaining unlabelled data points in the training set. Depending on the amount of labelled data present in the training set, the types of classification approaches and decision

rules obtained in semi-supervised learning can vary to resemble those obtained in the both supervised and unsupervised learning cases. Examples of semi-supervised learning include negative support vector data description (N-SVDD) [110] and cluster-based active learning [19].

2.7.2 Kernel Density Estimation

The kernel density estimate (KDE) is a non-parametric statistical approach to estimate the probability density of a variable [31]. The non-parametric nature of KDE makes it well suited to applications in which the underlying distribution of the data is not known a-priori, and provides a flexible means for PDF estimation in ML methods where the PDF is pre-requisite. Consider a variable $X = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^d$. Mathematically, the KDE of X is given by

$$p(X) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{X - x_i}{h}\right). \tag{2.29}$$

where $K(X; h)$ is a kernel function with kernel bandwidth h. Common choices for $K(X; h)$ in KDE include the uniform kernel and the Gaussian kernel. The uniform kernel is analogous to previously described Dirichlet window function, and the Gaussian kernel is given by

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{2.30}$$

The choice of the bandwidth parameter h has a strong influence on the clustering result. In supervised learning approaches, the bandwidth parameter h can be found using brute-force algorithms such as cross-validation [17].

2.7.3 Support Vector Machine (SVM)

SVM is a widely used supervised learning approach for classification [194]. The goal of SVM is to find a the hyperplane providing the maximum margin of separation between the separate two classes within a given a data set. This separating hyperplane can be sparsely represented by only a subset of points in the data set, which are known as the support vectors (SVs). Given a two-class data set $X = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^d$ with corresponding class labels $Y = \{y_1, y_2, ..., y_n\} \in \{-1, 1\}$, the separating hyperplane $f(X) = 0$ between these two classes is given by

$$f(X) = w^T X + b = \sum_{i=1}^{n} w^T x_i + b = 0 \tag{2.31}$$

where w represents a vector of weights and b is the bias. In order to define a separating hyperplane between the points in each class, each point in X should satisfy the follow constraints

$$\begin{aligned} w^T x_i + b \geq 1 \quad \text{if} \quad y = 1 \\ w^T x_i + b \leq -1 \quad \text{if} \quad y = -1 \end{aligned} \tag{2.32}$$

Equivalently, this can be expressed more succinctly by

$$y_i f(x_i) = y_i(w^T x_i + b) \geq 1 \qquad \forall \qquad i = 1, 2, ..., n \tag{2.33}$$

From Eq. 2.32, it can be shown that the only points within X required to fully define the separating hyperplane between to classes are those which satisfy the equalities $w^T x_i + b = 1$ and $w^T x_i + b \leq -1$. The points x_i satisfy these equalities represent the SVs. Oftentimes, a data set may contain classes which are not linearly separable in the original dimension $\mathbb{R}^d$. In these cases, a transformation is applied to the map the data into a higher-dimensional space (i.e. $\varphi : \mathbb{R}^d \to \mathbb{R}^{d'}$, where $d' \ggg d \in \mathbb{R}^{d'}$) where it becomes linearly separable [17].

2.7.4 Principle Component Analysis (PCA)

Principle component analysis (PCA) is a data dimensionality reduction technique widely used within statistical analysis and ML frameworks [50]. In practice, the two main purposes of dimensionality reduction are: for feature selection; and, to avoid the 'curse of dimensionality' [17]. For feature selection, dimensionality reduction is typically applied to remove non-informative, redundant (i.e. highly correlated) features from a feature set. Discarding these extra features enables a classifier to learn more efficiently and learn better decision rules using only the 'useful' features in the original feature set [160]. Within unsupervised learning, the most widely used dimensionality reduction technique is PCA [17].

Given a feature space of dimension D, the goal in PCA is to compute a transformation from D to a new set of coordinates (i.e. principal components) of dimension d (s.t. $d \lll D$), while retaining the maximum amount of variance from the original feature space. By definition then, the principal components (PCs) $U = \{u_1, u_2, ..., u_d\}$ form an orthonormal basis for the data. The PCs are arranged in order of decreasing variance (i.e. PC u_1 contains the largest amount of variance). Hence, in practice, value of d is determined by the desired level of variance to be retained after the transformation. Given a data set $X = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^D$, let u_i represent the projection of x_i in the direction of w such that

$$u_i = w^T x_i \qquad w^T w = 1 \tag{2.34}$$

The variance of U $Var(U)$ is given by

$$Var(U) = w^T X(w^T X)^T = w^T S w \qquad S = XX^T \tag{2.35}$$

Then, using the definition above, the goal of PCA is to determine the set w which maximizes $w^T S w$. The previously assigned constraint on the length of w i.e. $w^T w = 1$ ensures that the optimal solution for w is with respect to it's direction, rather than length. Eq. 2.35 can be re-written as the eigenvalue decomposition

$$w^T S w = \lambda \tag{2.36}$$

where w and λ are the eigenvectors and eigenvalues of S, respectively. Hence, the value w_1 corresponding to the largest eigenvalue λ_1 is the principal component u_1, and so forth.

2.7.5 Bayesian Updating

Bayesian updating is a statistical framework which uses Bayes' theorem to incrementally update a set of prior beliefs as more observations become available [185]. This property makes Bayesian updating particularly useful for applications where accurate prior knowledge is not available, as it allows for the discovery of the unknown prior distribution purely through observation. Given a model M described by set of parameters θ, Bayes' theorem states that the posterior probability of θ given X (i.e. $P(\theta|X)$) is given by

$$P(\theta|X) = \frac{\pi(\theta)P(X|\theta)}{P(X)} \tag{2.37}$$

where $\pi(\theta)$ is the prior probability, $P(X|\theta)$ is the likelihood, and $P(X)$ is the marginal probability. Since $P(X)$ is independent of the choice of θ, Bayes' rule is often expressed as the following proportionality instead

$$P(\theta|X) \propto \pi(\theta)P(X|\theta) \tag{2.38}$$

Let $\pi_1(\theta)$ represent the prior probability at some time step $t = 1$ where new observations become available. Bayesian updating states that $\pi_1(\theta)$ is equal to the posterior probability $P_1(X|\theta)$ computed at $t = 1$. In other words, the principle of Bayesian updating is to take the posterior of the previous update as the prior for the current update. Generalizing this for any time step $t = n$, the updated prior probability $\pi_n(\theta)$ is given by the hierarchical Bayesian framework

$$\pi_n(\theta) = P_n(\theta|X) \propto P_{n-1}(\theta|X)P_n(X|\theta) \tag{2.39}$$

Given enough observations of X over time, the framework in Eq. 2.39 will converge to the true distribution of $\pi(\theta)$ even if the initial estimate of $\pi(\theta)$ is poor, or non-informative [52].

2.8 Degradation Modeling and RUL Estimation

Degradation modeling is a central aspect of CBM prognostics which deals with the characterization and prediction of an observed degradation process and estimation of the RUL [128, 134]. The RUL of an asset is typically defined as the time remaining until the failure of that asset [61]. In statistical degradation modeling [128], the RUL is calculated as the time remaining until the prediction the modeled degradation process exceeds a specified failure threshold D_Y. In practice, D_Y is typically set using prior information such as historical failure data, or observed faults from similar units [102, 61]. Hence, for assets without this type of prior knowledge available, determination of a suitable failure threshold is a challenging task [102].

While both physics-based and data-driven approaches for degradation modeling [61] have been explored in the literature, the background presented in this section pertains specifically to the data-driven, parametric, random variable (i.e. statistical) degradation models employed in this thesis. Random variable (RV) degradation models assume that an underlying degradation process Y can be described using some pre-defined functional form η, subject to random effects β which may vary unit-to-unit. The general formulation for the RV model is given by

$$Y = \eta(t, \theta, \beta) + \epsilon \tag{2.40}$$

where θ are the parameters of functional form η, and ϵ represents the additive measurement error, which is typically assumed to be independent, zero-mean, and normally distributed with variance σ_ϵ^2. β is typically assumed to be multi-variate normal with mean μ_β and covariance Σ_β [128]. Commonly used functional forms of η include linear, exponential [52] or logistic [128].

One of the key practical benefits of parametric RV models is their ability to be initialized without explicit prior knowledge [52]. In these circumstances, the true model parameters can be iteratively refined using a suitable parameter updating framework such as the previously described hierarchical Bayesian framework. This property of parametric RV models makes them well suited for implementation in the unsupervised setting. Fig. 2.14 illustrates an example of a exponential RV degradation model using the hierarchical Bayesian framework for parameter updating. The degradation paths and model posteriors $P(\theta|Y)$ are shown at t = 2 years and t = 3.25 years. In addition, labels for the failure threshold D_Y and corresponding RUL are provided for the set of plots at t = 2. Comparison of the width between the posterior distributions obtained at the two time steps demonstrates the convergence towards the 'true' model parameters over time using Bayesian updating. A detailed derivation and justification of the exponential RV degradation model used in this thesis (Chapter 7) is presented in the corresponding methodology found in Section 7.3.9.

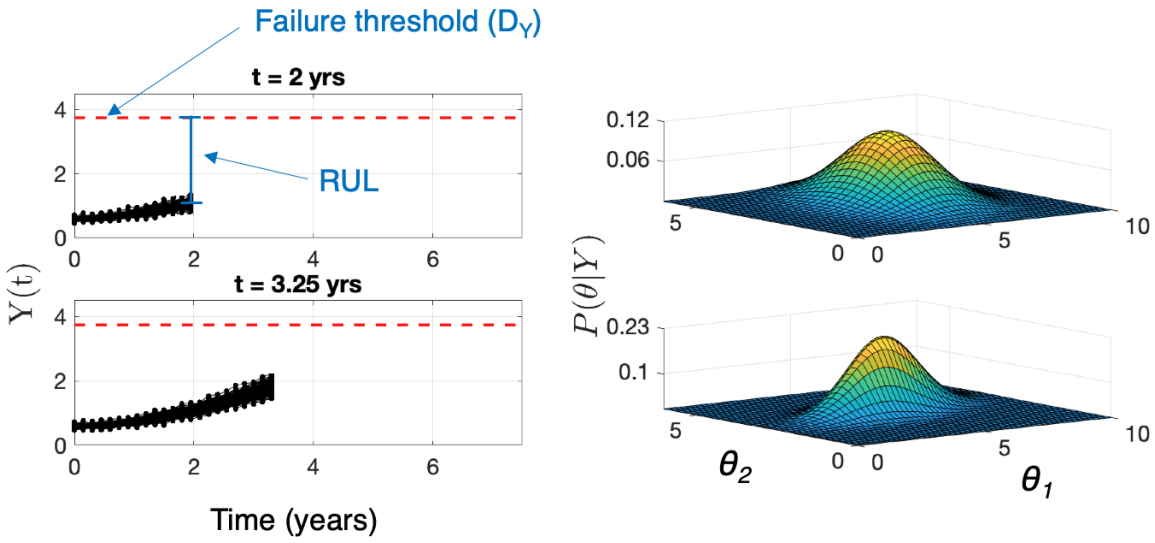

Figure 2.14: Example of exponential RV degradation model using hierarchical Bayesian updating. Degradation paths, failure threshold D_Y and posterior probabilities shown at t = 2 years (top) and t = 3.25 years (bottom)

2.9 Fault Vibration Signatures

Faults in rotating machinery can manifest across a number of critical components including gears [184], shafts [152] and bearings [153]. The range of possible fault conditions is further expanded by considering that each of these components can also experience a number of different types of faults ranging from misalignments, chips, cracks and spalls, to only name a few [152, 184]. The nuances and challenges specific to different types of faults has led to extensive bodies of fault-specific literature [104], which makes the development of an all-encompassing diagnosis or classification approach infeasible, if not impossible. However, due to the unsupervised nature of the present research problem, the underlying goal in this context is to simply detect a fault, rather than to diagnose or classify it. Hence, while classification of a particular fault may require special knowledge or analysis to reveal information *specific* to that fault—detection of faults at a more holistic level can be achieved by designing around characteristics *common* to a wide range of faults.

The generalizations applied towards machinery vibration signals to facilitate the development of robust detection methods towards the present research problem are as follows:

1. Vibration signals are comprised of both deterministic and stochastic components, alongside multiple noises processes. Common deterministic components include shaft frequencies, gearmeshing frequencies, and sidebands resulting from transmission errors [125, 152].

2. Gear faults will manifest in the form of amplitude and frequency modulations within the gearmesh harmonics [152]

3. Bearing fault signatures, which typically manifest as faint impulses in the vibration signal, are buried within the noise component of the signal [153, 136].

4. The appearance of a fault will induce changes in the composition of the signal assumed above which can be characterized using well-studied features extracted from the vibration signal [184, 59, 104]. These extracted features can be used to detect, but not diagnose the fault in the unsupervised setting.

5. Potential sources of non-stationarity in the signal will also induce changes in the signal composition. A detailed discussion on the types of on-stationarities central to the present research problem are presented in Chapter 4.

In reference to (4), examples of the typical types of features extracted from vibration signals used in this thesis for fault detection are listed below. References next to each type of feature are examples of their application to fault detection, while detailed descriptions of typical vibration features can be found in [59, 184, 104].

- Measures of Energy: (i.e. root mean square (RMS))[143, 105]: $x_{rms} = (\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{0.5}$
- Standardized Moments: (i.e. skewness and kurtosis) [143, 189, 147]: $\frac{\mathbb{E}[(X-\mu)^n]}{\sigma^n}$
- Measures of Peakiness: (i.e. crest factor) [36, 11]: $CF = \frac{max|x_i|}{x_{rms}}$
- Measures of Entropy (i.e. Shannon entropy)[59]: $H(x) = \mathbb{E}[-\log(P(X))]$
- Equivalent features to the above in the Fourier domain (i.e. mean frequency, spectral entropy) [6, 103, 110, 143].

2.10 Summary

This chapter presents the overarching theoretical background used in this thesis, beginning with a comprehensive overview of the CBM architecture and it's constituent components. Following this, the fundamental concepts within signal processing, machine learning, statistics and degradation modeling are presented. The chapter concludes by presenting key generalizations of gearbox vibration signatures used within the proposed methodology.

Chapter 3

Literature Review

This chapter presents a review of CBM literature relevant to the research problem and proposed methodology, beginning with an overview of the main research areas and trends in CBM to help contextualize the specific motivations behind the present body of work. Following this, the state-of-the-art from three research areas directly related to the present body of work are discussed in the context of unsupervised CBM. These research areas are: pre-processing and vibration analysis methods for non-stationary vibration signals, data-driven fault detection/condition monitoring; and, data-driven degradation modeling. A holistic summary of the key gaps across all three research areas is provided at the end of the chapter.

3.1 Core Areas and Research Trends in CBM

The body of literature underneath the umbrella of CBM spans across a wide range of research fields including maintenance policy, machine learning [104], signal processing [51], diagnostics [35] and prognostics [61], with active contributions from multiple disciplines ranging from engineering to decision sciences. A recent review by Quatrini et al. [148] analyzed over 4000 contributions attempted to quantify the main research areas and research trends within CBM over the last 30 years. Their review identified the following four key research areas within CBM: CBM fundamentals, CBM strategies (encompassing fault detection and condition monitoring), unsupervised methods, and prognostics. CBM fundamentals primarily encompasses high-level contributions discussing the role of CBM in maintenance, policy and planning [3, 40, 43]. Apart from an initial spike in research

activity in the early 2000s coinciding with the mainstream inception of CBM as a maintenance strategy, research interest in CBM fundamentals has remained relatively constant over the past two decades [148].

At the next level down, CBM strategies encompasses the broadest range of topics, ranging from specific signal processing, fault detection and condition monitoring strategies [121, 200, 84] to detailed frameworks for implementation [71, 18, 142, 178]. Relative to CBM fundamentals, these contributions approach CBM at a more technical level. Research activity in CBM strategies gained significant traction between 2008 and 2015, but has decreased over the last few years [148]. The rise and fall of research activity in this category is perhaps correlated to the increasing presence of engineering systems of comprised of multiple, interconnected systems, which demand analytical capabilities and robust frameworks beyond what is currently available in the literature [104]. A key distinction to make here is that the drop in contribution rate in recent years is not for lack of interest, but rather an increase in the complexity and novelty of the problem to be solved [104]. At the same time, the ubiquity of sensing in modern engineering systems and increasing maturity of ML literature in recent years [87] has paved the way for the development of new, data-driven and ML-based analytical approaches [61, 104]. Additional evidence to support these claims can found by examining the underlying trends within prognostic research over the last decade [148]. Apart from the general rise in popularity, much of the prognostic research in recent years has shifted towards use of data-driven ML approaches for feature learning, health and RUL estimation [204, 87, 85].

The final key research area identified in their review pertains to unsupervised CBM methods, and specifically, CBM methods which rely solely on the sensor data. Research activity in unsupervised methods remained relatively silent between 2008 and 2015 [148]. The lack of interest over this period can be attributed to the same reasons previously used to explain the trajectory of prognostics research. Namely, prior to 2015, there was no real demand for unsupervised CBM methods, as the existing tools within literature were more than adequate to address the needs at the time [59, 86]. The key difference between unsupervised methods and prognostics however (and the reason why prognostics achieved traction much earlier on), is that the need for prognostics in CBM was there since the inception of CBM [181]. In recent years, the same needs that drive the development of data-driven ML approaches in prognostics have motivated new research activity within unsupervised methods as well [104]. In response to the increasing complexity of engineering systems, recent contributions from unsupervised methods [67, 129, 169] and semi-supervised methods [124, 110, 104] are also focused on the development of tools to enable the implementation of CBM in applications where prior knowledge (i.e. operating characteristics or failure records) is not available, to enable the implementation of CBM

in applications where the system characteristics and/or operational behaviour cannot be easily understood. These central motivations are also the driving mechanism behind the body of work presented in this thesis.

3.1.1 Specific Research Areas

With respect to the research trends and motivations presented in Section 3.1, the body of work presented in this thesis is also focused on the development of new, unsupervised and non-parametric tools for CBM in response to the novel research problems posed by modern day engineering systems. Specifically, the proposed methodology aims to tackle specific knowledge gaps in relation to the availability of non-parametric pre-processing and vibration analysis of non-stationary engineering systems; and, unsupervised tools for data-driven condition fault detection, thresholding and degradation modeling. Hence, the literature review presented hereafter will focus on the following three specific research areas within CBM:

1. Pre-processing and vibration analysis methods for non-stationary vibration signals
2. Data-driven fault detection
3. Data-driven degradation modeling

The following sections will present the state-of-the-art in each of these areas to identify the key knowledge gaps with respect the unsupervised CBM to be addressed by the proposed methodology.

Limitations in Scope

The limitations in the scope of the literature presented in this chapter are as follows:

1. Emphasis on Vibration-based Approaches — while CBM has been studied throughout literature using a number of different sensor measurements, including acoustic [73], temperature [66], current [14], oil debris [41] measurements, the vast majority of CBM studies are conducted using vibration measurements [153] due to their high sensitivity to machinery faults [186, 86] and relative ease to which they can be obtained. The present body of work is also formulated around vibration data, and hence, the literature presented in this chapter will pertain predominantly to vibration-based CBM methods.

2. Emphasis on Data-driven Approaches — while both fault detection and degradation modeling have been studied in CBM using model-based approaches [35, 82, 49], these approaches generally suffer from inflexibility and the inability to be updated using sensor data [61], and therefore are not well-suited to the present research problem. Hence, the literature presented and discussed in this chapter will focus primarily on data-driven approaches for fault detection and degradation modeling.

3.2 Pre-processing and Vibration Analysis

Realization of an effective unsupervised vibration-based CBM framework will require robust and non-parametric pre-processing and vibration analysis tools, that can address the unique challenges present in the unsupervised setting. Broadly speaking, the primary function of pre-processing is to maximize the visibility of desired signal components within a given signal [60]. In this sense, "pre-processing" is an application-agnostic term and ubiquitous across all fields of science and engineering that deal with signals or sensor data [170]. Vibration analysis on the other hand, has deep roots within CBM, and encompasses both pre-processing techniques [65, 138] and signal processing-based methods for fault detection and diagnosis [151, 11, 152] specific to vibration signals. Hence, the emphasis of the discussion presented in this section will be placed on the-state-of-the-art from the latter. This section begins by presenting an overview of the state-of-the-art in vibration analysis and pre-processing in CBM to help contextualize the detailed discussions that follow.

3.2.1 Vibration Analysis

In the realm of CBM and machinery monitoring, vibration analysis has been demonstrated to be a powerful tool for fault detection and condition monitoring [184, 59, 1]. When a fault develops within a mechanical component, the information pertaining to the fault will often manifest found the vibration signal. This information can be extracted using a wide variety of pre-processing and vibration analysis techniques, of which the nature and complexity will vary depending on the application. Examples of simple approaches include time-domain, feature-based approaches which involve the extraction and monitoring of signal statistics such as RMS, crest factor and kurtosis [78, 59], and basic Fourier analysis (i.e. DFT [170]), which can be applied to stationary signals to identify fault frequencies in the spectrum. More sophisticated techniques, which are typically used for the analysis of non-stationary signals, will typically approach the problem in the time-frequency domain [65, 137, 9, 70] or angular domain [64, 15, 8]. Examples of these techniques include cepstral

editing [138, 151], minimum entropy deconvolution (MED) [26], cyclo-stationary analysis [23], and time synchronous averaging [15], to name a few. The common tradeoff found in many current techniques is that increased analytical capability or robustness comes at the expense of increased computational complexity and heavy dependency on prior knowledge of the operational characteristics or system kinematics [184]. This is especially true for the case of non-stationary signals.

Vibration analysis techniques for non-stationary signals will typically require some form pre-processing as a pre-requisite. In these techniques, the role of pre-processing is to either directly enhance the presence of diagnostic information buried within the signal [65, 137], or to convert the signal into a form that is more suitable for extracting diagnostic information [64, 15, 138]. Depending on the type of fault, the critical diagnostic information can be found within different components of the signal. As discussed in Section 2.9, gear fault signatures will typically manifest within the harmonic content of the signal, while bearing faults can manifest themselves in both the harmonic or stochastic components of a signal [152]. Hence, a common goal in both pre-processing approaches is the separation of the deterministic and stochastic components of a signal.

3.2.2 Pre-processing Techniques for Non-Stationary Vibration Signals

Time-frequency-based pre-processing approaches such as the Fourier synchrosqueezing transform (FSST), MED [26] [2, 137], sliding window singular spectrum analysis [65], automated cepstral editing (ACEP) [138, 151], and stochastic resonance method [136] have all been applied to the spectral analysis of non-stationary signals. While highly effective, these methods are computationally expensive, and require user expertise for parameter tuning, filter design, as well as prior knowledge in order to be effective. Hence, these methods are not conducive to online monitoring or the unsupervised setting. Similarly, angular domain-based approaches such as auto-regressive TSA [11] and multi-order probabilistic approach [8] add an additional layer of computational complexity and constraints to the problem by requiring the instantaneous shaft speed to either be measured or estimated using computationally expensive algorithms [34, 99, 8, 30], which makes these types of approaches difficult to implement at scale or in the unsupervised setting. Vibration analysis in the unsupervised setting requires the use of autonomous, unsupervised or non-parametric approaches.

3.2.3 Autonomous Vibration Analysis Techniques for Non-Stationary Vibration Signals

In the context of autonomous vibration analysis techniques for non-stationary signals, literature is relatively sparse and generally inapplicable to non-stationary signals containing time-varying harmonic components. Schoukens et al. proposed a method for autonomous spectral analysis [165]. However, the proposed method was limited to analysis of periodic signals, and cannot be applied to highly non-stationary signals. Barbe and Van Moer presented another autonomous approach for the detection of peaks buried within noise using discriminant analysis in conjunction with thresholds [13]. However, this method yields optimal results only if the underlying Gaussian assumption is satisfied. Additionally, the thresholds used in their study were heuristically determined. Martin et al. proposed an automatic spectral analysis technique using a multi-estimator approach that can be used to autonomously extract and classify spectral peaks in a stationary vibration signal [125]. The proposed method is parametric and the thresholds pertaining to the spectral estimators are user-defined and dependent on the probability distribution of the data, which may vary significantly in non-stationary signals. Singh et al. proposed a real-time pre-processing approach for non-stationary signals based on the orthogonal empirical mode decomposition (OMED) [159, 117] The use of OMED over EMD provides practical advantage of increased robustness to noise. However, like EMD, the performance of OMED is heavily influenced by the choice of frequency resolution. Under non-ideal conditions (i.e. low sampling frequency), EMD approaches are unable to separate low-amplitude harmonic components in signals containing dominant harmonic frequencies [113]. Though not explicitly stated in their contribution, Ma et al. proposed a matching synchroextracting transform (MSET) approach based on the synchroextracting transform (SET) and instantaneous frequency for extraction of time-varying harmonics in non-stationary signals [122], which can be adapted to autonomous applications. However, the results obtained MSET are heavily predicated on the quality of the initial estimate obtained for the fundamental instantaneous frequency, which may be unreliable in highly non-stationary or low-SNR conditions.

Apart from autonomous pre-processing and vibration analysis techniques required for the spectral analysis of non-stationary signals, dedicated pre-processing tools to address other forms of potential non-stationary present dynamic systems are also required for successful implementation. For example, in condition monitoring of machinery with multiple discrete operating states (i.e. multimodal systems), the differences in the operational behaviour within each state will manifest in the vibration signal as well. The unsupervised problem will require autonomous pre-processing tools that are capable of differentiating vibration data between each operating state without knowledge of the operating charac-

teristics. Outside of vibration analysis, autonomous pre-processing techniques applicable to this problem can be found in clustering literature [156]. In the context of CBM, clustering methods such as K-means [17], Hidden Markov Models (HMMs) [55] and Gaussian Mixture Models (GMMs) [155] have all been applied to fault diagnosis [95, 74, 75, 135]. Nelwamondo et al. [135] applied HMM and GMM clustering in conjunction with mel-frequency cepstrum coefficients for bearing fault diagnosis. Heyns et al [75] applied the negative log likelihood of GMM from windowed TSA segments as a feature for fault detection in non-stationary conditions. A subsequent study by Heyns et al. [75] proposed an updated GMM-based fault detection approach for non-stationary systems which does not require the use of TSA. Schmidt et al. [164] proposed a fault diagnosis approach for non-stationary signals using an HMM-based discrepancy criteria. In all of these approaches, the role of clustering is to facilitate the diagnosis or direct detection of an underlying fault within a non-stationary signal, which is a significantly more complex (i.e. computationally expensive) solution than the present problem (i.e. unsupervised CBM) requires. The present problem requires a computationally efficient tool for unsupervised classification of the current machine operating state.

Pertaining more towards operating state classification, Zhang et al. [206] proposed an approach for transient and steady-state discrimination for bearings using variational-Bayesian GMMs (v-GMM). In their research problem, only one operating state is known initially. Detection of the unknown transient operating state was performed by monitoring for changes in the vGMM clustering result as new data is added to the model. This, in essence, is still anomaly detection as opposed to unsupervised classification, since the framework used in their study does not leverage the inference aspect of GMMs. Such is also the case in the approach proposed by Cao et al. [22], which uses the Kullback-Leibler (KL) divergence of the entire GMM result to perform holistic condition monitoring of a multimodal manufacturing process. The works presented by Zhang et al. and Cao et al. represent complete, but focused solutions to specific CBM problems. The present problem requires a more flexible clustering tool that can be easily implemented alongside other pre-processing tools in the unsupervised setting. Such types of problems have been investigated within other areas of research [88, 206, 37], but have yet to be fully explored in the context of unsupervised CBM.

3.3 Data-Driven Fault Detection

Fault detection is primary means deployed in condition monitoring to identify anomalies (i.e. faults) in real-time [188]. For complex systems of systems, where degradation be-

haviour may be difficult to predict, fault detection and condition monitoring are quintessential to maintaining the desired level of reliability during operation. Within the hierarchy of CBM diagnostics, fault detection occurs at the first layer, preceding fault isolation and fault diagnosis [152]. Hence, fault detection is solely concerned with obtaining a binary classification result. Detailed analysis, such as determining the location, nature and cause of a fault are functionalities reserved for fault isolation and fault diagnosis [104]. As the desired level of detail in diagnostic information increases, so does the complexity of the analysis. In general, fault isolation and fault classification either require domain expertise (i.e. methods described in Sections 3.2.1 and 3.2.2) and/or specific prior knowledge of the system (i.e. fault frequencies, system kinematics, labelled data), which increases the difficulty of real-time implementation [51, 104]. To draw a healthy balance between reliability, computational complexity and deployability, most unsupervised CM and CBM frameworks will aim to implement fault detection only.

This section presents the discussion on the relevant literature within the state-of-the-art of supervised, semi-supervised and unsupervised fault detection and condition monitoring literature. Definitions of these classifications were first introduced in Section 2.7.1. This section begins by presenting the past and present research trajectory of supervised fault detection, which will be essential for contextualizing the motivations and contributions within semi-supervised and unsupervised fault detection. Given the nature of supervised learning (i.e. the use of labelled data), fault detection and classification will occur simultaneously in most of these approaches. Hence, where appropriate, the term "classification" will be used in place of "detection".

3.3.1 Supervised Fault Detection

Early supervised fault detection approaches, which began to gain traction near the turn of the century, employed statistical or traditional ML tools for fault detection [104]. Between 2000 and 2015, the body of work in supervised fault detection was dominated by statistical approaches such as support vector networks (SVNs) [33], probabilistic graphical models [94], NNs [199] and SVMs [194] as classification tools. The general methodology used in traditional, supervised data-driven fault detection is to pre-process and extract handcrafted, fault-sensitive features from labelled data, which are then used to train a classifier to perform fault classification [104, 190]. Typical features include time-domain features such as RMS, kurtosis, crest factor, mean frequency, and standard deviation frequency, to name a few [103]. During this period, significant research interest was directed towards SVM-based approaches [79] due to their strong performance on small datasets, high interpretability over black-box approaches such as NNs, and high consistency in training as

a result of convex optimization [103]. These properties have helped to maintain a consistent research interest in SVM-based fault detection approaches even up to present day [141, 108].

During the advent of deep learning and big data, research endeavours turned towards the use of deep-learning (DL) architectures for feature learning and fault detection [104]. The main types of DL tools used in the literature include stacked de-noising autoencoders (SDAEs) [158], generative adversarial networks (GANs) [57], deep belief networks (DBNs) [76], and convolutional neural networks (CNNs) [96]. In particular, hierarchical feature learning approaches, which employ DL methods such as stacked autoencoders (AEs) to learn features from latent, non-linear relationships in data [63, 87, 197], have gained significant traction over the last few years. When these features are applied in similar DL architectures for fault classification [83, 69, 24, 119], state-of-the-art approaches have already demonstrated near perfect classification accuracy on most publicly available datasets [209]. Hierarchical feature learning - while powerful - requires both significant user expertise and domain knowledge in order to design and tune the network architecture to achieve reasonable results. An important point to make here is that hierarchical feature learning can be applied in both the supervised [63, 87, 197] and unsupervised setting [67, 129]. The caveat is that the features extracted using these black-box approaches suffer from low interpretability compared to traditional, handcrafted features, of which their relationships to different types of machinery faults are well-understood [59, 103]. In the unsupervised setting, handcrafted features can offer intrinsic, interpretable diagnostic value in the absence of all other information that hierarchical features cannot.

Given the rapid advancement and performance saturation of supervised fault classification approaches over the last half decade, recent research interest has shifted towards addressing the practical shortcomings of these approaches. Namely, the strong, if not absolute dependency on labelled data for detection and classification means that these approaches cannot be easily applied to novel systems or applications in which adequate training data is not available (i.e. the present research problem) [104]. Existing supervised fault classification approaches require not only a sufficient volume of labelled data, but also a relatively balanced dataset - the latter of which is extremely difficult to obtain in most real-world applications. Secondly, the heavily parametric nature and computational complexity of DL-based approaches is a major barrier towards implementation in real-time online monitoring, unsupervised settings and large-scale applications [104]. These practical considerations have propelled diagnostic research along two divergent pathways. The first path involves the use of transfer learning theory [140] to enable the implementation of existing supervised diagnostic approaches in novel applications and scenarios with class imbalance. The principle of transfer learning is to facilitate the application of knowledge

learned in one "known" setting to a new, but related setting (e.g. two different types of machinery with some similarity). While several exploratory studies have already been published on the topic [198, 146, 192, 168], the research area of transfer learning is still in it's infancy, and hence - there are significant milestones left to achieve before transfer learning approaches can be applied efficiently in practice. The second path forward involves the use of semi-supervised [110, 124] and unsupervised approaches [67, 130, 169] for fault detection. Due to the relaxed nature of their learning requirements relative to supervised approaches, semi-supervised and unsupervised fault detection approaches are inherently more conducive for implementation on novel systems, online monitoring and large scale applications.

3.3.2 Semi-Supervised Fault Detection

Situated between supervised and unsupervised approaches, semi-supervised fault detection approaches encompass a wide range of methods and frameworks, including contributions which lean closer to supervised learning [19] and others which border on unsupervised learning [124]. Close evaluation of semi-supervised works is crucial even when designing supervised and unsupervised approaches, as many semi-supervised approaches are inspired by practical constraints and limitations encountered during the application of particular supervised and unsupervised methods. For example, the practical benefits of applying a particular supervised approach to an application containing only partially labelled data may motivate the modification of that supervised approach to work under the new set of constraints. Relative to supervised fault detection approaches, the relaxed requirements of semi-supervised learning makes this class of approaches more conducive to real-time monitoring.

In comparison to the body work surrounding supervised fault detection approaches, the literature pertaining to unsupervised and semi-supervised methods is relatively sparse. Semi-supervised approaches were proposed by Liu et al. [110], Mao et al. [124], Bull et al. [19] and Potočnik et al. [143]. Liu et al. [110] proposed a semi-supervised fault detection method for rolling element bearings based on features obtained from cyclic spectral analysis and a negative support vector data description (NSVDD) classifier. To train their NSVDD, the uniform object generation method was used to augment the unlabelled training set with a small fraction of labelled, synthetic anomalous data. Additionally, a new multi-criteria fault detection threshold is introduced, in which the severity of an anomaly is based on the number of criteria met. This multi-criteria threshold introduces the idea of differentiating between the beginning of the degradation process and a fault, but the proposed threshold cannot be used for smooth prediction since the classification levels are based on binary

criteria. Mao et al. [124] proposed an online semi-supervised fault detection framework using SDAE feature learning and a safe semi-supervised support vector machine (S4VM). A limitation of this approach is that a small fraction of labelled faulty samples are required for training, which means it cannot be applied in scenarios where only healthy data is available. Additionally, the hyperparameter tuning approach is not explicitly clear, and the detection threshold is determined a-posteriori. Bull et. al [19] proposed a semi-supervised approach for vibration-based structural damage detection on an aircraft wing using cluster-based active learning. The principle of active learning is to reduce the amount of labelled data required for learning in big data applications, but assumes that the corresponding labels for a given datapoint can be queried if need [167]. Hence, the main contribution of their approach is a primarily a method for optimized supervised learning, rather than a dedicated semi-supervised detection framework. Potočnik et al. [143] presented a semi-supervised framework for multi-class classification of compressor operating state using handcrafted features extracted from unlabelled data. Class labels for a small subset of the training data were inferred using a heuristic measure. The study compared the classification results obtained using several types of classifiers including linear discriminant analysis (LDA), NNs, extreme learning machines (ELMs) and SVMs. Among the classifiers considered, non-linear SVM was identified as a robust choice, given it's consistency in training (i.e. convex optimization) and strong ability to generalize. The main challenge identified with SVM is the proper selection of the hyperparameters.

3.3.3 Unsupervised Fault Detection

Unlike supervised and even semi-supervised approaches, the complete lack of prior information available in the unsupervised setting poses a unique set of challenges for fault detection. Historically (see Section 3.1), these challenges and limitations have functioned as a deterrent to detailed exploration of unsupervised approaches so long as feasible solutions existed within supervised and unsupervised learning [148]. Recently however, the novel problems introduced by the increasing complexity of modern day engineering systems - along with the practical constraints inhibiting the implementation of existing supervised and semi-supervised approaches to these problems, has renewed research interest in unsupervised fault detection approaches [104, 157, 72]. The common framework used in unsupervised fault detection is to evaluate fault detection as a one-class problem. In other words, the assumption or underlying inductive bias [115] in unsupervised learning is that the data observed pertains only to the negative (i.e. healthy) class. A primary challenge then in unsupervised fault detection is the trade-off between detection sensitivity and generalization performance i.e. how does one decide how far out to set the decision boundary between

classes if they only have knowledge of one of those classes? This paradox commonly manifests itself in the area of parameter estimation in unsupervised approaches [169, 39, 130], as well as in threshold estimation in data-driven prognostics [61] and is a knowledge gap that has yet to be sufficiently addressed in the literature.

Fully unsupervised CBM approaches were proposed by Hasani et al. [67], Michau and Fink [129], Shi et al. [169], Diez-Olivan et al. [39]. Hasani et al. [67] proposed an unsupervised fault detection framework based on the correlation rate of features extracted using an AE. In their study, the degradation starting point is defined as the time the auto-encoder correlation rate falls below 90% of its value from $t-100$ steps prior. Although fully unsupervised, the use of vanilla AE architecture for feature learning results in significant training times, and the fault threshold used is determined heuristically. Michau and Fink [129] presented a fully unsupervised approach using a stacked extreme learning machine auto encoder (ELM-AE). The ELM architecture drastically reduces the training time compared to vanilla AE. Michau and Fink also proposed that the distance of an anomalous point from the healthy class can be used to infer the severity of a fault. The main limitations of their approach are that the fault detection threshold is set heuristically, and that the ELM-AE hyperparameters are tuned based on maximizing the detection accuracy a-posteriori. Shi et al. [169] proposed an unsupervised fault detection approach using a stacked de-noising auto encoder (SDAE) and long short-term memory (LSTM) architecture for feature learning. The SDAE reconstruction error was used to train a Support Vector Data Description (SVDD) classifier for fault detection. However, the process used to estimate the SVDD hyperparameters is not clear. Like Michau and Fink, Shi et. al. also proposes the concept of using the classifier distance as a measure of severity. The concept of distance-based severity inference was also proposed by Diez-Olivan et. al [39] in their kernel-density estimate (KDE) based One Class Support Vector Machine (OC-SVM) approach to fault detection. In [39], sigmoid normalization is applied to the classifier output to normalize the results of multiple OC-SVM classifiers. The use of KDE in the hyperparameter estimation process introduces an additional bandwidth parameter that must be optimized.

3.4 Data-Driven Degradation Modeling

Prognostics and degradation modeling play a fundamental role within the CBM framework [98]. Accurate estimates of prognostic parameters, such as the RUL, can provide operators with invaluable information to optimize maintenance actions, operational planning and replacement scheduling. This is particularly true for systems with uncharacterized degra-

dation or failure behaviour, for which a deeper understanding of these characteristics can only be drawn from the data. As shown in Section 3.1, research interest in prognostics, especially data-driven prognostics has continued to grow over the last decade - partially in response to the increasing complexity of the systems to be modeled, but also as a result of the emergence of new tools in AI and ML which help address existing challenges within prognostics such as operational uncertainties [207]. The challenges posed by operational uncertainties are further compounded in the unsupervised case, where the current body of literature remains relatively unexplored [61, 207, 116]. While the methodology (i.e. fault prediction) proposed in this thesis is not concerned specifically with the traditional prognostic framework (i.e. RUL/failure estimation), a discussion on the state-of-the-art in traditional data-driven prognostics is still necessary to understand the knowledge gaps and motivations behind the proposed approach. This section presents a high-level review of literature pertaining to statistical and ML-based degradation modeling, and identifies the overarching knowledge gaps which motivate the proposed methodology.

3.4.1 Statistical Degradation Modeling

Similar to the dichotomy found in data-driven fault detection, data-driven approaches in prognostics can also be divided into statistical and ML-based approaches [61, 207]. Statistical approaches include regression-based models [120, 144, 90, 52], hidden Markov models (HMMs) [161, 127], proportional hazard models (PHMs) [180, 205], and Bayesian networks (BNs) [126, 116].

Regression-based models operate on the assumption that the underlying degradation trend, captured through a CI or set of CIs, can be mapped to some functional form [120]. Within statistical degradation modeling, regression-based models are the most commonly employed statistical degradation modeling approach, due to their relative simplicity and high interpretability [61]. The mathematical convenience of certain models, such as the exponential degradation model [52], makes these models computationally efficient and suitable for real-time applications. A key barrier to their implementation in unsupervised frameworks is that the required failure threshold is determined using prior knowledge (i.e. failure records) [40, 97, 109, 102, 61]. The topic of threshold estimation in the absence of prior knowledge has not been a topic of focus yet in prognostics literature [61]. Additionally, without a proper framework in place for updating prior beliefs [27], regression-based models may perform poorly under highly uncertainty conditions.

HMM-based approaches offer a powerful means to map the relationship between CIs and the actual underlying degradation process, but are limited in their predictive capabilities

due to the underlying Markovian assumption and lack of closed form solution for the RUL [61]. HMMs also suffer from the same pitfall as regression-based models, in that the failure threshold is typically defined using prior knowledge [61]. PHM approaches [180, 205] offer the advantage of high model intepretability. However, the absolute requirement for failure or censored data makes them unconducive to unsupervised applications [127, 61]. BNs, and specifically dynamic Bayesian networks [116], combine the Markovian principle with graph theory and the reliability concept of the P-F curve [131] to model degradation as a function of time-invariant processes and observations [81]. Due to the their inherent ability to use observations to influence prior beliefs, dynamic BNs are well equipped to handle and respond to uncertainty [116]. Furthermore, the use of the P-F principle means that a degradation threshold does not need to be defined explicitly using prior knowledge [131]. However, the construction of the BN architecture from data in BNs and dynamic BNs is largely dependent upon domain knowledge and user expertise [182], which makes BNs unconducive to unsupervised applications.

3.4.2 ML-based Degradation Modeling

ML-based approaches for degradation modeling have continued to increase in popularity over the last decade [61, 207]. Powerful characteristics of ML tools, such as the ability to learn latent relationships between observed variables [199, 63, 87], and the ability to construct model-free representations from the data [194], have propositioned ML as an attractive solution to both new and existing challenges within data-driven prognostics [61, 207]. Examples of ML applications in prognostics include fuzzy logic (FL) condition state estimation [106, 212], NN-based RUL estimation [177, 53, 28], SVM-based regression for RUL estimation [80, 118, 171, 180], and DL tools such as RNNs [112], CNNs [42] and LSTMs [210] have also been applied to RUL estimation.

Among these tools, SVM has been the most widely applied ML tool in data-driven degradation modeling [80]. The strong generalization performance of SVM means that it is both effective and straightforward to implement across a wide range of applications [80]. Tran et al. [180] applied SVM in conjunction with PHM for RUL estimation. However, the use of PHM means that failure data is required. Benkedjouh et. al [16] proposed the use of the SVDD decision boundary width as a CI for degradation modeling. In their approach, the failure threshold is defined directly from the data, using an assumption placed on the true distribution of the SVDD boundary width. However, the assumption used requires the data to be well-behaved and compact in the healthy state of the machine, as well as the degradation behaviour to be monotonic. The study by Benkedjouh et. al represents one of few work to explicitly address the problem of failure threshold estimation. Sun

et al. [171] proposed a Bayesian least squares SVM approach for RUL estimation. The Bayesian framework was applied to recursively update the SVM model parameters. The use of a Bayesian framework helped to improve the robustness of the RUL prediction under uncertainty. However, the failure threshold in this approach is still predefined.

The robustness of the Bayesian learning to types of the uncertainties present in degradation modeling, including model parameter estimation, has led to the exploration of relevance vector machines (RVMs) for RUL estimation [195, 111, 21, 38]. RVM is a sparse Bayesian SVM approach that provide a probabilistic classification result [179]. The use of Bayesian learning in RVM means that the SVM model parameters can be estimated using iterative Bayesian parameter estimation approaches (i.e. expectation-maximization (EM) [179]. Widodo et al. [195] presented an RVM approach for estimating survival probability using partially-censored data. While the primary motivation behind their work is to address the ubiqituous failure threshold paradox in RUL estimation, the proposed approach still requires the use some failure data, and hence, cannot be applied to applications containing no historical failures. Liu et al. [111] studied the performance of RVM for real-time RUL estimation on small datasets, and demonstrated that RVM - and by extension, SVM - are well-suited to real-time or online monitoring applications. Hybrid approaches, which aim to leverage the advantages of RVM and statistical regression models such as the exponential degradation model [38] and logistic degradation model [21], have been demonstrated within the literature to be strong performers as well. While the use of Bayesian learning in RVM alleviates the issue of parameter estimation, the use of the EM algorithm means that computational efficiency of RVMs scale poorly with increasing size of data, and convergence to the global maximum is not guaranteed [179].

3.5 Summary of Research Gaps

The increasing complexity of modern engineering systems and limited applicability of existing supervised and semi-supervised CBM approaches to these applications has created a new demand for the development of CBM solutions conducive to these systems. Within CBM literature, the unique requirements of this novel research problem has motivated solution finding along two divergent pathways. Within supervised CBM, the response has come in the form of increased research interest in transfer learning methodologies. At the same time, the congruent nature of semi-supervised and unsupervised frameworks to the specific challenges presented has led to increased research interest in the development of solutions within both of these areas. While interest in semi-supervised/unsupervised frameworks has expanded, many of the tools required to facilitate the realization of such

frameworks in practice have yet to be proposed in the literature. In relation to unsupervised CBM, the key knowledge gaps within the current-state-of-the-art - which are to be addressed by the proposed methodology - are as follows:

1. In relation to pre-processing and vibration analysis:

   (a) The majority of existing pre-processing and vibration analysis approaches for non-stationary signals require significant user expertise, heavy parameter tuning and prior knowledge (i.e. angular measurements, operating characteristics, system kinematics) to be effective. These factors inhibit the deployment of these tools in the unsupervised setting, real-time monitoring and large scale applications.

   (b) Existing methods for autonomous spectral analysis are generally not applicable to non-stationary signals.

   (c) Non-stationarities present in modern engineering systems (i.e. systems of systems) include systemic, stochastic and multimodal sources of non-stationarity. In conjunction with autonomous tools for non-stationary spectral analysis, the dynamic operational behaviour of such engineering systems will require additional, unsupervised pre-processing tools (i.e. clustering methods) to further differentiate between different operating conditions in the unsupervised setting. While such methods have been applied within CBM in the context of fault diagnosis, the application of such methods—especially in conjunction with other unsupervised pre-processing tools—has yet to be explored in detail within the context of unsupervised CBM.

2. In relation to fault detection:

   (a) The strong generalization performance of SVM-based classifiers has lead to extensive application of such methods within unsupervised fault detection. However, the problem of hyperparameter estimation, particularly in the unsupervised setting, has yet to be properly addressed within the literature.

   (b) The majority of existing unsupervised and semi-supervised fault detection approaches apply black-box hierarchical feature learning to extract CIs from the data. Unlike traditional handcrafted features - of which their relationships to various types of machinery faults is well understood - the interpretability of black-box features is low.

3. In relation to unsupervised degradation modeling:

(a) Determination of a suitable degradation threshold is fundamental for RUL estimation. In almost all cases, however, the degradation threshold is either predefined (i.e. heuristically set) or derived from prior knowledge such as historical failure data. Without a means for determining a suitable threshold in the absence of prior knowledge, RUL estimates cannot be obtained in the unsupervised setting. Apart from a handful of studies, the problem of prior-less degradation thresholding has yet to be explored in the literature.

(b) Existing unsupervised approaches are heavily focused on the detection aspect of CBM. Additional exploration of dedicated, unsupervised prognostic or degradation modeling methods are required to achieve a complete implementation of CBM in the unsupervised setting.

Chapter 4 will present a detailed description of the specific research problem and proposed methodology to address each of the aforementioned gap areas.

Chapter 4

Problem Description and Methodology

Chapter 3 identified key gaps in literature across the various fields of research relevant to unsupervised CBM. The purpose of this chapter is to present the various methodologies proposed in this thesis to tackle each of these gap areas, and contextualize these contributions within the overall unsupervised CBM framework. In other words, this Chapter is meant to act as a map to illustrate the connections between the various methodologies within this body of work.

This chapter is structured as follows. Section 4.2 presents a high level description of the unsupervised CBM framework and the relationship of each gap area with respect to the overall framework. Section 4.2.1 provides a summary of these gap areas along with the proposed methodology to tackle each of the knowledge gaps. Lastly, Section 4.4 provides an overview of the two field pilots at Toronto Pearson International Airport and additional data sets used for validation in this thesis.

4.1 Types of Non-Stationarity in Machinery Vibration Signals

Non-stationarities in machinery vibration signals originate from a number of physical phenomena and operational parameters. Classifying the types of non-stationarity types examined in this thesis using traditional statistical definitions (i.e. first order stationary, weak-sense stationary) described in Section 2.3.2 may not provide the clearest or most

intuitive description of these processes. Alternatively, the types of non-stationarity encountered in this thesis can be categorized with respect to the physical phenomena that causes them, with real examples provided to further illustrate each case of non-stationary behaviour. This Section presents the main categorizations of non-stationarity addressed in this thesis and provides context in relation to the types challenges posed by each type of non-stationarity towards unsupervised CBM.

4.1.1 Non-Stationarity Arising From Multimodal Operation

The first categorization pertains to non-stationarity found in multimodal machinery i.e. machinery with multiple, discrete operating states. The differences in the operating characteristics between each operating state will manifest accordingly in the vibration signal, and in it's entirety, the resulting signal will be non-stationary. However, the stipulating assumption applied in this thesis is that the vibration behaviour within a given operating state is stationary. Hence, if pre-processing can be applied to a multi-operating state signal to decompose the signal into a set of constituent, stationary signals (where each signal pertains to one operating state), then conventional vibration analysis techniques for stationary analysis can be easily applied to the resulting signals.

A clear example of multimodal non-stationarity is found in the Pearson Terminal 1 passenger boarding tunnel pre-conditioned air (PCA) units. Each PCA unit has a finite set of pre-configured settings to service different sizes of aircraft. The differences between each operating states are clearly manifested in the vibration behaviour of the unit. The typical vibration signal and corresponding sliding mean frequency for a PCA unit with four discrete settings is shown in Fig. 4.1. The vibration signal in Fig. 4.1 is comprised of three discrete phases, each associated with a specific level of variance, which suggests the presence of three discrete operating states. However, closer evaluation of the corresponding sliding mean frequency of the vibration signal clearly illustrates the presence of four, well-separated operating states, each described by a different mean frequency.

4.1.2 Systemic and Stochastic Non-Stationarity

The second categorization of non-stationarity broadly encompasses all non-stationarities resulting from the time-varying and stochastic processes present in the machinery. The most common examples of these non-stationarities are speed and load fluctuations - which can occur deterministically (i.e. systematic speed regulation via control system) or stochastically (i.e wind loading on a wind turbine). Several key challenges arise in the analysis of

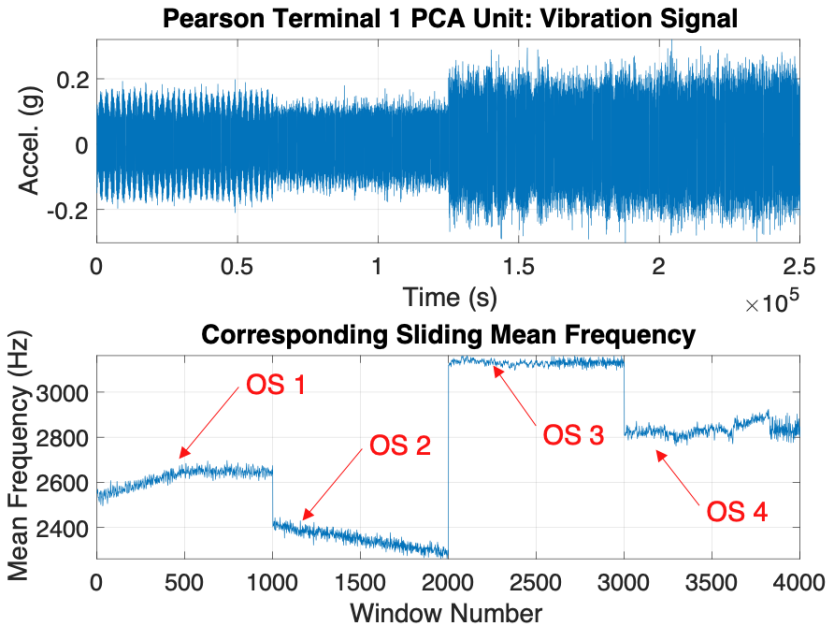

Figure 4.1: Example of multi-operating state non-stationarity in Pearson Terminal 1 PCA units: vibration signal (top) and corresponding sliding mean frequency (bottom)

signals containing these types of non-stationarity. For vibration analysis and fault detection approaches that rely upon the tracking of harmonic content, speed and load fluctuations will induce frequency and amplitude modulations that will result in sub-optimal conditions for analysis. In the supervised setting, the common solution is to apply domain expertise and time-frequency analysis methods to pre-process the signals.

Fig. 4.2 provides an example of stochastic non-stationarity found in the wind turbine vibration data obtained from the CMMNO' 14 [29] challenge. The amplitude modulation of the signal is clear in the time domain, and the corresponding spectrogram helps to illustrate the resulting frequency modulation caused by the stochastic wind loading.

4.1.3 Combination Non-Stationarity

The two categorizations of non-stationarity presented are not mutually exclusive. While only a subset of real-world examples will exhibit multimodal non-stationarity, almost all real-world examples will contain some degree of systemic and stochastic non-stationarity.

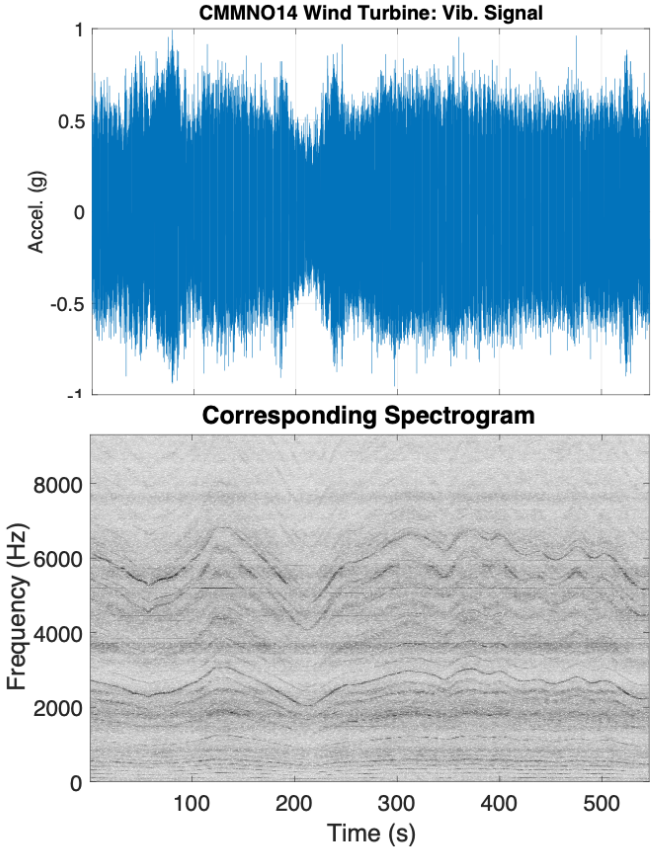

Figure 4.2: Example of systemic and stochastic non-stationary found in CMMNO'14 wind turbine signal: vibration signal (top) and corresponding spectrogram (bottom)

However, in many cases, the latter type of non-stationarity may not severe enough to require sophisticated pre-processing to be applied to the vibration signal prior to analysis. An example of a real-world system in which both categorizations of non-stationarity are non-negligible is the LINK Pearson APM gearbox. Fig. 4.3 illustrates the vibration signal and corresponding spectrogram of the APM gearbox for one complete cycle of the train's travel. The APM vibration signal contains clear examples of discrete operating states, stochastic non-stationarity in the form of short-duration impulse-like phenomena, as well as systemic non-stationarity during the train's acceleration and deceleration phases, which result in the time-varying harmonics shown in the corresponding spectrogram. Hence, pre-processing of such signals will require a combination of methods to tackle both types of

non-stationarity.

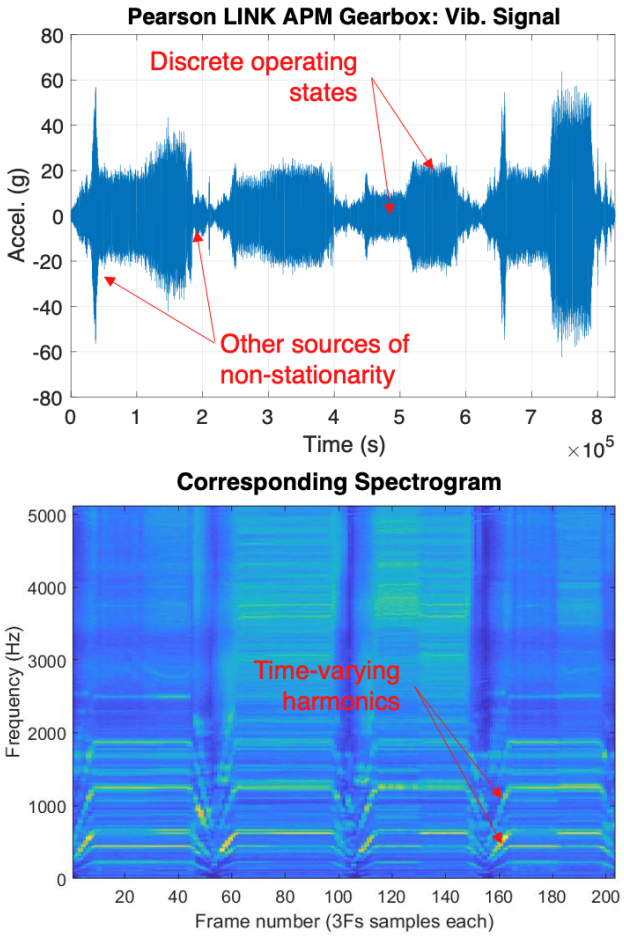

Figure 4.3: Example of combination non-stationarity in Pearson LINK APM Gearbox: vibration signal (top) and corresponding spectrogram (bottom)

4.2 Problem Description

Relative to the general CBM structure presented in Section 2.1 and Fig. 2.1, the key differences between traditional (i.e. supervised) CBM and unsupervised CBM stem from the complete absence of prior knowledge in the latter. As shown in Section 3.2, many

(if not all) existing pre-processing and vibration analysis techniques for non-stationary signals require some form of prior knowledge in order to be effective. Taking into account practical considerations as well, the majority of existing methods presented in Section 3.2.2 are computationally expensive and/or require significant user expertise, which makes them unconducive to online monitoring or large-scale CBM applications.

4.2.1 Key Gap Areas in Unsupervised CBM

Fig. 4.4 presents the high-level structure of an unsupervised CBM framework. Key gap areas are denoted by the red dotted lines, while the corresponding labels indicate the proposed solution methodologies, which will be presented in the following Section. Fig. 4.4 helps to visualize the key disconnects within the framework predominantly exist within the pre-processing level prognostic levels. The main observations from Fig. 4.4 are as follows:

1. Without the development of specific unsupervised approaches for pre-processing the various types of non-stationary data found in machinery vibration signals, current unsupervised CBM frameworks are not able to deliver the high-quality pre-processed data necessary for subsequent analyses. The specific approaches required are:
   (a) A computationally efficient and robust non-parametric pre-processing method for general (i.e. systemic and stochastic) non-stationarity (Section 4.1.2), which can be using for de-noising, and extraction of time-varying harmonic components for use in subsequent vibration analyses. The proposed solution for this is S-MSC (Chapter 5).
   (b) An unsupervised means for blind decomposition of a multimodal vibration signal (Section 4.1.1) into a set of stationary, constituent signals to which common vibration analysis methods can be applied. The proposed solution for this is GMM-OSD (Chapter 6).
2. Pertaining to unsupervised fault detection and degradation modeling, the main knowledge gaps pertain to parameter estimation. For fault detection, existing unsupervised hyperparameter estimation approaches for the one class problem perform poorly when applied to noisy real-world data [54, 114]. For degradation modeling, there are a lack of methods available to determine a suitable threshold for degradation modeling without the use of prior knowledge. Hence, the specific approaches required are:

(a) A fully unsupervised parameter estimation approach for hyperparameter estimation in one-class fault detection applications.

(b) A fully prior-independent approach for determining a suitable failure threshold for degradation modeling.

The methodology developed in this thesis to address both these gaps is the B-OCSVM (Chapter 7). The following Section presents a detailed methodology for the solutions developed in this thesis.

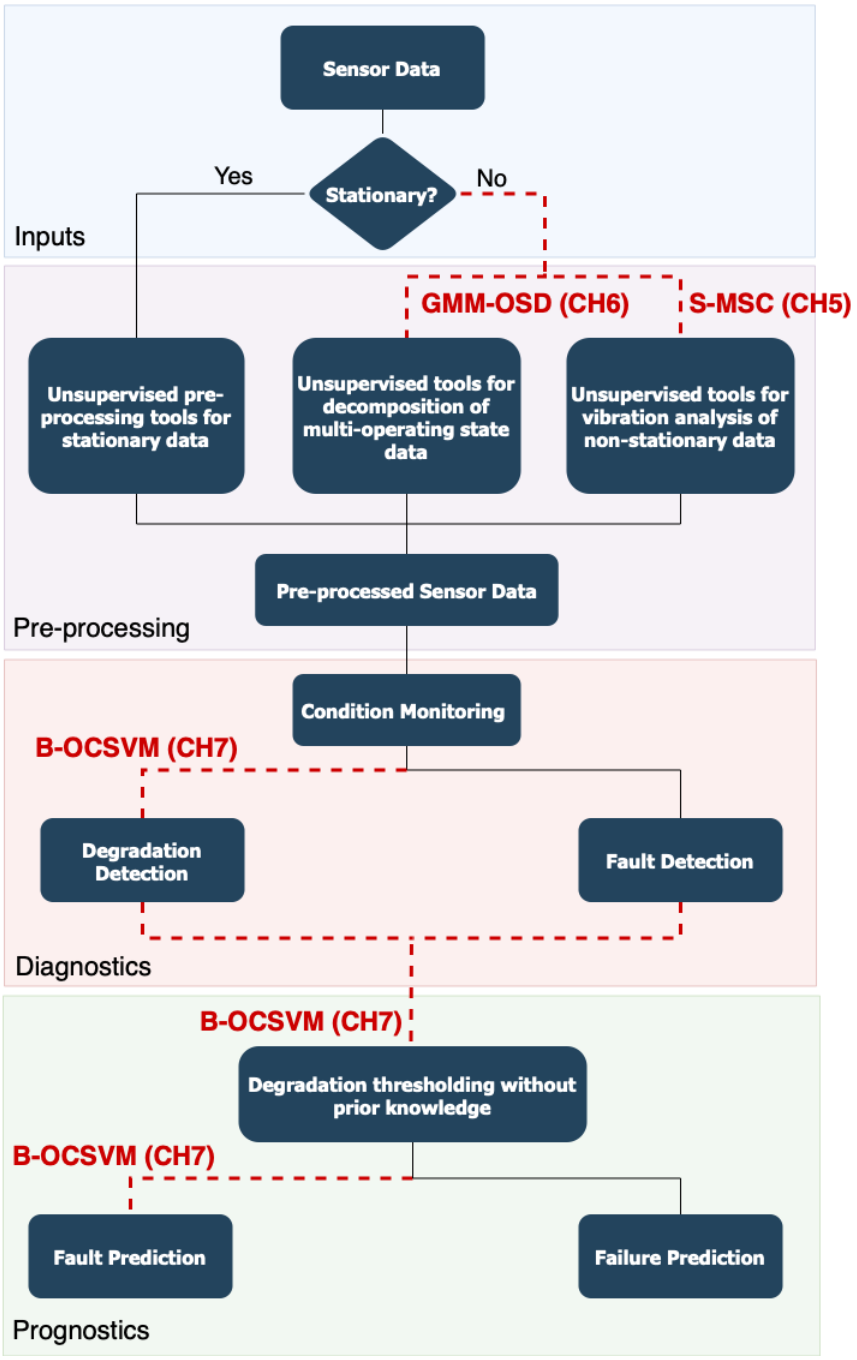

Figure 4.4: Overview of unsupervised CBM framework: red dotted lines denote gap areas and corresponding labels indicate proposed solution methodology and relevant thesis chapters

4.3 Methodology

The three methodologies presented in this thesis to tackle the existing gaps in unsupervised CBM are: Spectral Mean Shift Clustering (S-MSC), Gaussian Mixture Model Operating State Decomposition (GMM-OSD), and Bayesian One Class Support Vector Machine (B-OCSVM). The purpose of this Section is to provide a concise description of the methodology employed by each approach. To improve the clarity and organization of the work in presented in this thesis, the detailed background and descriptions of each approach can be found in their respective chapters. Descriptions of the datasets used to validate each approach are presented at the end of this Section.

4.3.1 Spectral Mean Shift Clustering (S-MSC)

The S-MSC approach is proposed as a robust, non-parametric time-frequency approach for pre-processing and spectral analysis of signals containing systemic and stochastic non-stationarity. The S-MSC approach applies mean shift clustering (MSC) [32] to the STFT power spectra to separate time-varying harmonic components from noise and other stochastic low-energy signal components. Section 5.8 presents an unsupervised approach to estimate the kernel bandwidth parameter in MSC. The validation approach for S-MSC is as follows:

1. De-noising and extraction of time-varying harmonic content from the LINK APM gearbox data (Section 5.5), including:
   (a) Comparison against the marginal spectrum
   (b) Comparison against Time Synchronous Averaging (TSA)
2. Pre-processing and envelope analysis for non-stationary bearing fault detection on the Safran aircraft engine dataset (Section 5.6)
3. Comparison against the following state-of-the-art time-frequency spectral analysis approaches using the CMMNO'14 wind turbine dataset (Section 5.7):
   (a) Singular Spectrum Analysis (SST)
   (b) Fourier Synchrosqueezing Transform (FSST)

The S-MSC approach is presented in Chapter 5.

4.3.2 GMM-OSD

The proposed GMM-OSD approach is a soft-clustering pre-processing approach for unsupervised decomposition of multi-state machinery vibration signals. GMM-OSD employs Gaussian Mixture Modeling (GMM) [135] in conjunction with CIs extracted from the vibration signal to decompose a multimodal vibration signal into stationary, constituent clusters - each representing a discrete operating state of the machine. Applied together with S-MSC, these two approaches form a robust pre-processing framework for the unsupervised analysis of non-stationary signals. OSD is presented using two GMM methods:

1. A classical GMM approach using Bayesian Information Criterion (BIC) for model selection
2. A variational Bayesian GMM approach

The validation approach for GMM-OSD is as follows:

1. Supervised study on the proposed feature set and featured-based operating state characterization of the LINK Pearson APM gearbox and Terminal 1 passenger boarding tunnel PCA unit data (Section 6.3)
2. Application of GMM-BIC approach (Section 6.4):
   (a) OSD of Terminal 1 PCA vibration data
   (b) OSD of LINK APM gearbox vibration data with no pre-processing applied
   (c) OSD of LINK APM gearbox vibration data with S-MSC pre-processing
   (d) Application of the vGMM approach (Section 6.5):
      i. OSD of Terminal 1 PCA vibration data
      ii. OSD of LINK APM gearbox vibration data with S-MSC pre-processing

The GMM-OSD approach is presented in Chapter 6.

4.3.3 B-OCSVM

The proposed B-OCSVM approach is an unsupervised approach for early degradation detection and fault prediction building upon OC-SVM [166] and the hierarchical Bayesian

framework [52]. Unlike traditional fault detection approaches, B-OCSVM targets the detection of the degradation process before it manifests into a fault. In conjunction, B-OCSVM employs hierarchical Bayesian degradation modeling to predict the time at which a degradation process will manifest into a fault. For deployability in unsupervised CBM, new unsupervised methods are proposed for OC-SVM hyperparameter estimation (Section 7.3.5) and degradation model thresholding (Section 7.3.7). The validation approach for B-OCSVM is as follows:

1. Application of B-OCSVM to the IMS bearing dataset 7.6:
   (a) Benchmarking against current state-of-the-art ML fault detection approaches
   (b) Sensitivity analysis of the proposed OC-SVM hyperparamter estimation approach
2. Application of B-OCSVM to the FEMTO bearing dataset to validate the generalization performance across multiple operating states and multi-fault bearings (Section 7.7).
3. Application of B-OCSVM to the C-MAPSS turbofan engine dataset 7.8:
   (a) Validation of generalization performance to non-bearing faults
   (b) Validation of generalization performance on sparsely sampled data

The B-OCSVM approach is presented in Chapter 7.

4.3.4 Practical Considerations

In addition to technical requirements, two additional practical considerations are central to the design of each approach. The first design consideration is the ability of said approach to integrate into IoT frameworks. IoT readiness plays a crucial role in expediting the deployment of these methods into real world applications. The second consideration is the ability of these approaches to integrate cleanly into existing asset management frameworks already in use within organizations, to minimize disruptions to workflow and reduce the number of barriers to implementation. In other words, the secondary overarching objective is to make the proposed approaches easily implementable in real world applications.

4.4 Field Implementations

Vibration data was collected from two field pilots located at Toronto Pearson International Airport. This Section presents the details of both studies.

4.4.1 Pearson Link APM Gearbox

The Pearson LINK APM train shown in Fig. 4.5a is a cable-driven APM system that provides an around-the-clock transportation service for passengers and employees between the airport's two terminals (Terminal 1 and 3) and adjacent parking facilities (Fig. 4.5b. The trains themselves are computer-controlled, cable-driven carriages that are devoid of any drive assemblies. Power is generated from a central station that houses all of the drive train machinery, and tractive forces are delivered to the train via a cable connecting the gearbox to the underside of the carriage. Hence, the LINK APM gearbox was determined to be the critical component to be monitored during the field pilot.

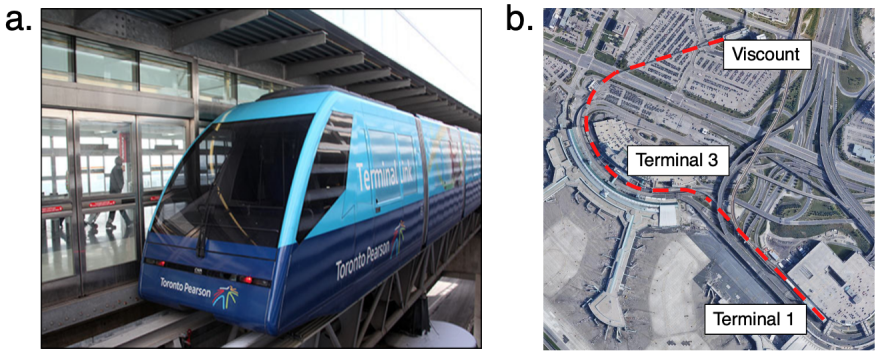

Figure 4.5: Pearson LINK APM a) Train b) Route

Gearbox Description

The LINK gearbox (Fig. 4.6 is a three-stage gearbox with a maximum input shaft speed of 1340 rpm and corresponding maximum output shaft speed of 54 rpm. The fundamental gear mesh frequencies (h1) and corresponding harmonics for the gearbox are shown in Table 4.1 for each gearbox stage. Each shaft of the gearbox was instrumented with a uniaxial PCB model: 352C68 accelerometer (A1, A2, A3, A4 in Fig. 4.7), measuring

tangential acceleration. The gearbox and accelerometer mounting positions are illustrated in Fig 4.7. Accelerometer data is collected using a Siemens LMS SCADAS Mobile (LMS) data acquisition system. Continuous vibration data was sampled at 20 kHz during normal operation of the train. Each test corresponds to one complete round trip of the train over the route shown in Fig. 4.5b.

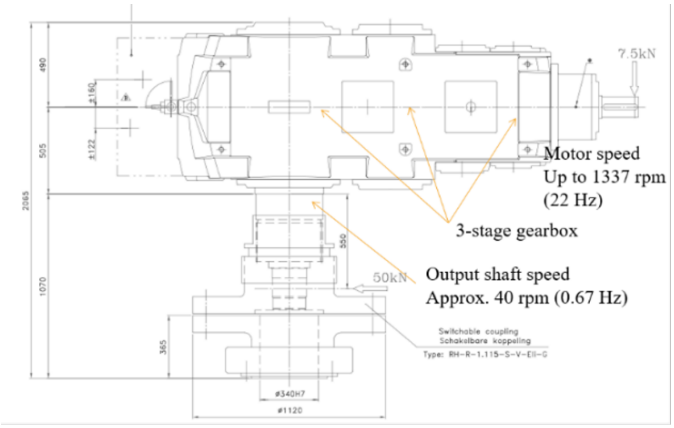

Figure 4.6: LINK APM gearbox: plan view

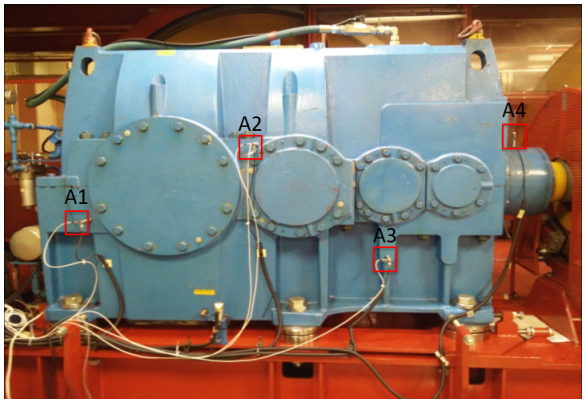

Figure 4.7: LINK APM gearbox: accelerometer mounting locations

Collection of Angular Data for Time Synchronous Averaging

In addition to the collection of vibration data, angular measurements (i.e. order tracking) of the input and output shaft were collected using a Monarch PLT 200 Tachometer. To fa-

Gearbox Stage	h_1	h_2	h_3	h_4
a (Input)	624	1248	1872	2496
b (Intermediate)	221.8	443.6	665.4	887.2
c (Output)	47.23	94.5	141.7	189

Table 4.1: LINK APM gearbox stages: gearmeshing frequencies (h_1) and harmonics (Hz)

cilitate order tracking, the input and output shafts of the the APM gearbox were retrofitted with reflective taping as shown in Fig. 4.8. For tests including angular measurements, vibration data was sampled at 200 kHZ to provide adequate frequency resolution during synchronous averaging.

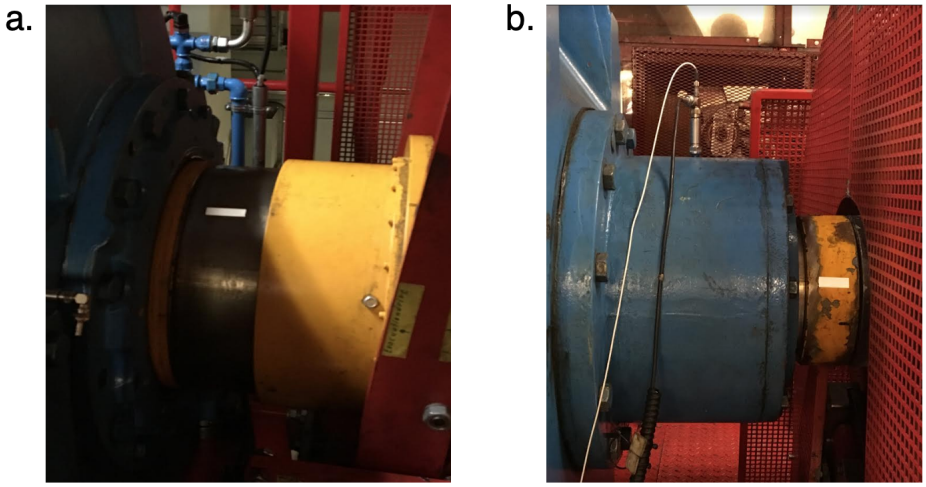

Figure 4.8: LINK APM gearbox - angular measurement locations for: a) Output shaft b) Input Shaft

4.4.2 Pearson Terminal 1 PCA Units

The second, ongoing field pilot at Toronto Pearson International Airport is conducted on a set of pre-conditioned air (PCA) units housed on passenger boarding tunnels in Terminal 1. The scope of work consists of a preliminary study to collect vibration data from each unit to characterize the vibration behaviour of each PCA unit, followed by a long-term

IoT CBM study. The preliminary phase of work was completed in 2021, and the pilot is currently in the implementation process for the IoT CBM study. The study area is shown in Fig. 4.9. The PCA units are used to filter, heat, cool and remove moisture from air before circulating it into the cabins of aircraft while they are docked at a gate for passenger boarding. The PCA units under study operate on 4 discrete settings, dependant on the size of aircraft docked at the gate. The typical exterior, interior and sensor mounting location for a PCA unit are shown in Fig. 4.10 a, b and c, respectively.

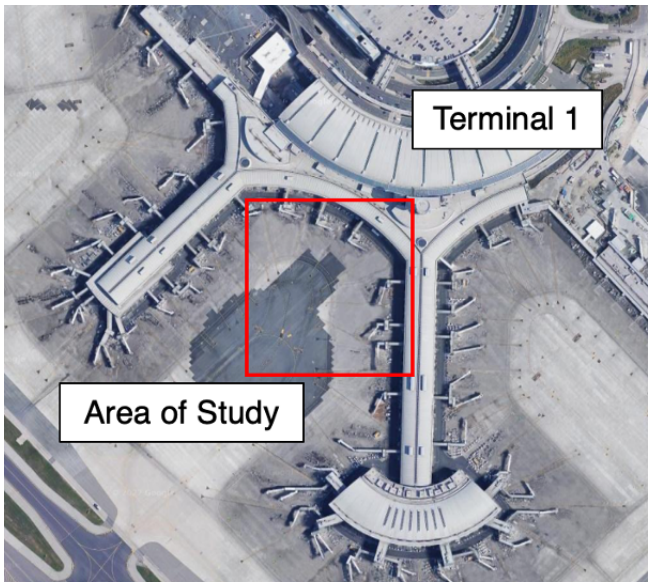

Figure 4.9: Pearson Terminal 1 PCA Field Pilot - Area of Study

Preliminary Study

For the preliminary study, the main motor bearing housing of each unit was instrumented with a uniaxial accelerometer (PCB model: 352C68) in the location shown in Fig. 4.10c. Vibration data in the tangential direction, sampled at 20kHz, was collected using was collected using the LMS data acquisition system. For each unit, 2-5 minutes of continuous vibration data was collected for each operating state.

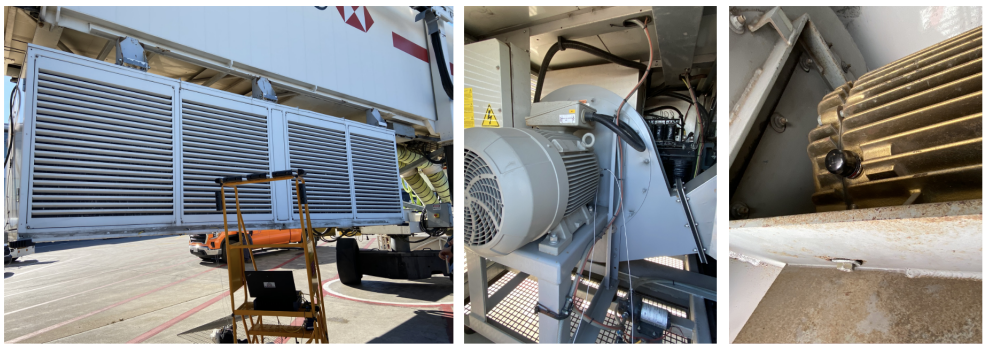

Figure 4.10: Typical Pearson terminal 1 passenger boarding tunnel PCA unit: (a) exterior, (b) interior (c) sensor location on motor bearing housing

4.5 Additional Datasets

While the datasets obtained from the field pilots at Toronto Pearson International Airport contain valuable examples of non-stationarity in real-world systems, the lack of observed failure and degradation behaviour in the data requires the use of additional datasets for validation of the proposed approaches. The following datasets are publicly available machinery vibration datasets, which contain fault or run-to-failure data suitable for validation of the proposed approaches. The information presented here is simply to provide a high-level description of the unique properties of each dataset. Detailed descriptions of each dataset will be presented as they appear in validation.

1. **CMMNO'14 Wind Turbine Dataset:** data from wind turbine containing roller bearing fault. Signals containing significant, non-linear speed fluctuations and other sources of stochastic non-stationarity (Section 5.7).

2. **Surveillance 8 Challenge - Safran Engine Dataset:** data from Safran aircraft engine containing two faulty bearings. Signals collected during linear acceleration of the engine, resulting in time-varying bearing fault frequencies (Section 5.6).

3. **IMS Run-to-Failure Bearing Dataset:** run-to-failure bearing data. Widely used for benchmarking in ML-based fault detection (Section 7.5.1).

4. **FEMTO Run-to-Failure Bearing Dataset:** run-to-failure bearing data collected across multiple operating states. Failed bearings are presumed to contain every type of bearing fault (Section 7.5.2).

5. **C-MAPSS Run-to-Failure Turbofan Engine Dataset:** simulated run-to-failure turbofan engine data. Data is sampled at a very low frequency (1Hz) (Section 7.5.3).

4.6 Summary

This chapter presents the detailed description of the research problem and the overall methodology to address the key knowledge gaps identified in Chapter 3. High-level descriptions for each component of the overall methodology are presented, followed by descriptions of the two field pilots located at Toronto Pearson International Airport.

Chapter 5

Spectral Mean Shift Clustering (S-MSC)

This chapter presents a non-parametric pre-processing tool for analysis of non-stationary signals, termed spectral mean shift clustering (S-MSC). The purpose of S-MSC is to provide a robust, pre-processing tool that can be used for analysis of non-stationary vibration signals in the absence of prior knowledge. S-MSC tackles the following key gaps in literature - namely, the lack of computationally efficient, blind pre-processing methods conducive to online CM frameworks; and, the lack of non-parametric pre-processing methods for spectral analysis of non-stationary vibration signals.

The chapter begins by summarizing the specific gaps in knowledge and motivations behind the approach, followed by section 5.3, which presents the background and detailed description of the methodology. Sections 5.4-5.7 presents validation results obtained through different applications of the proposed approach, including several use cases for the approach in vibration analysis, application to different types of machinery signals, and evaluation of the performance of S-MSC against state-of-the-art spectral pre-processing approaches. Following validation, section 5.8 proposes an automated estimator for the S-MSC kernel bandwidth parameter.

5.1 Motivations for S-MSC

While a detailed discussion on the key knowledge gaps within the current state-of-the-art for pre-processing of non-stationary vibration signals is presented in 3.2, the following

summary is intended to provide the reader with the specific context required for understanding the motivations behind the work presented hereafter. The key observations from the current-state-of-the art are as follows:

1. The main purpose of pre-processing in machinery diagnostics is to either directly enhance the presence of diagnostic information buried within the signal, or to convert the signal into a form that is more suitable for extracting diagnostic information. Depending on the type of fault, the critical diagnostic information can be found within different components of the signal: gear faults typically manifest themselves within the harmonic components, while bearing faults can manifest themselves within the harmonic or stochastic components of a signal [152]. Hence, the separation of the deterministic and stochastic components of a signal is a fundamental pre-processing phase to many diagnostic approaches.

2. The majority of pre-processing techniques suitable for the analysis non-stationary vibration signals require significant domain knowledge or iterative parameter tuning in order to be effective. In addition, many of these methods are computationally expensive or depend upon other pieces of prior knowledge (i.e. angular reference signals), which further inhibits the deployability of these approaches in online CM and CBM applications

3. Spectral (i.e. time frequency) methods are particularly effective for analysis non-stationary signals. A central focus in many of these spectral tools is the extraction of time-varying harmonic components, or in a more general sense, to enhance the visibility of diagnostic information buried in a time-varying signal. However, most if not all existing methods for this are parametric, and hence, cannot be easily applied to applications where the required model parameters are not known a-priori. These methods also suffer from the same aforementioned practical limitations makes them unconducive for online CM and CBM applications.

5.2 Background

S-MSC is a spectral pre-processing method for machinery diagnostics, which can be used to separate time-varying harmonics from noise or stochastic components in non-stationary vibration signals. The proposed approach aims to address the aforementioned knowledge gaps in 5.1 by adopting a computationally efficient, non-parametric approach for pre-processing and spectral analysis of non-stationary signals that can be easily implemented

within online CM and CBM applications. The proposed approach uses MSC in conjunction with the STFT to blindly separate the noise and time-varying harmonic components within each spectrum. One key property of the proposed approach is that the signal decomposition behaves like a linear filter in that the properties of the extracted components of the signal (harmonics and noise), including phase and magnitude, are preserved after performing separation using MSC. This means that the original signal can be reconstructed simply combining the IFTs of each decomposition component. Hence, the proposed approach can be used as a pre-processing technique for many existing fault detection methods.

5.2.1 Mean Shift Clustering

MSC is a clustering algorithm that performs density mode detection on the KDE of a data [32]. The MSC algorithm consists of two sequential steps: KDE estimation followed by hill climbing algorithm to find then the local maxima (i.e. clusters) of the KDE [32]. MSC is a non-parametric algorithm, meaning that it does not require any a-priori knowledge or assumptions to be made regarding the underlying distribution of the data. The non-parametric nature of MSC also implies that the probability of convergence is initialization-independent, in contrast to other popular clustering techniques such as Gaussian Mixtures or K-Means [123]. These properties make MSC well suited to the unsupervised setting and for large-scale CM and CBM applications [25]. Prior to the work presented in this chapter, the application of MSC as a time-frequency analysis tool for machinery diagnostics has not previously been studied.

5.2.2 Spectral Mean Shift Clustering

The approach used in this work, termed S-MSC is blind time-frequency signal processing technique for non-stationary gearbox vibration signals. The principle of S-MSC is to apply MSC to the power spectral amplitudes obtained from each STFT window in a given vibration signal to separate the time-varying harmonic components from noise components. In the context of an arbitrary gearbox vibration signal, S-MSC blindly separates a signal into two components: a de-noised signal containing only the time-varying harmonics and other deterministic frequencies, and a residual component containing the noise.

The signal decomposition of S-MSC is best understood through an example. Consider the composition of a typical gearbox vibration spectrum shown in Fig. 5.1a. The frequency content in the spectrum can be broadly categorized into two groups: a group consisting of sparse higher energy components or peaks, corresponding to harmonics and other dominant

processes in the gearbox, and a second group containing dense lower energy broadband noise and/or non-dominant gearbox processes. Examples of these two groups are shown in the STFT spectrum of the LINK APM gearbox in Fig. 5.1a. Consider also the 1-D representation of the spectrum in Fig. 5.1a shown in Fig. 5.1b, which is created by flattening the frequency axis. This flattened spectrum is the input to S-MSC, except that in practice, no labels (i.e. noise/peaks) are assigned to the frequency components a-priori. The KDE of the flattened spectrum shown in Fig. 5.1c represents the starting point for the S-MSC clustering algorithm. Fig. 5.1c illustrates that the peak corresponding to the density of the noise components is significantly higher than those corresponding to harmonics and other dominant gearbox frequencies. Hence, it can be assumed that in S-MSC that the highest density cluster corresponds to noise, thereby allowing it to be blindly separated from the signal. In other words, the density distribution, or rather the contrast in density between the two aforementioned groups of signal components in relation to one-another is what enables their blind separation using S-MSC.

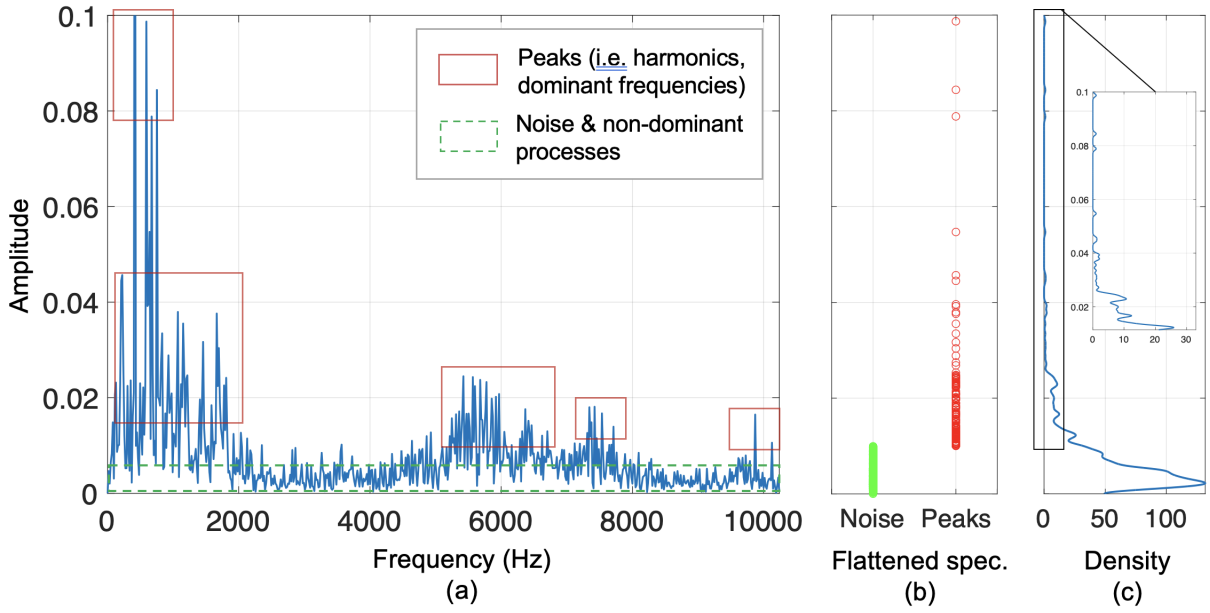

Figure 5.1: Typical composition of gearbox vibration spectrum: (a) LINK Pearson APM gearbox STFT spectrum, (b) frequency-flattened spectrum of (a), and (c) kernel density estimate of (a)

Unlike many existing model-based approaches or TSA-based techniques, a major advantage of S-MSC is that it is not dependent on priori knowledge. Spectral decomposition

using S-MSC can be performed directly on the STFT of the raw time signal, without the need for knowledge of gearbox kinematics or resampling in the angular domain using an angular reference signal as is often required in many current state of the art pre-processing techniques for non-stationary vibration signals. The ability to directly and simultaneously extract both the time-varying harmonic components and the noise from raw non-stationary signals in a single step significantly reduces the computational complexity, which is of particular practical importance for online CBM applications.

5.3 Methodology

This section presents the methodology for the S-MSC approach.

5.3.1 Basic Assumptions

Consider a discrete gearbox vibration signal $x[n]$ sampled from a continuous time process $x(t)$ such that

$$x[n] = x(\frac{n}{f_s}) \tag{5.1}$$

where, n is the sample number, and f_s is the sampling rate. It is assumed that $x(t)$ follows the basic assumptions described in 2.3.1 such that

$$x(t) = c(t) + e(t). \tag{5.2}$$

where $c(t)$ contains the harmonic content in the signal and $e(t)$ is an additive term comprised of noise and other stochastic processes.

5.3.2 Spectrogram Representation

Given a discrete signal $x[n]$ of length L satisfying the representation in Eq. 5.2, the STFT of $x[n]$ using the definition found in Eq. 2.20 is given as

$$\begin{aligned} X(m,k) = \sum_{n=0}^{N-1} x(n+mH)w[n]e^{-j2\pi kn/N} \\ M = \frac{L-N}{H} \\ m = 1, 2, ..., M \\ k = 1, 2, ..., K \end{aligned} \tag{5.3}$$

where m dictates the window number, k represents the k^{th} spectral component, $w[n]$ is a window function of length N, $K = N/2$ is the total number of spectral components as dictated by Nyquist theorem, H represents the shift length between two consecutive windows and M is the total number of windows. Taking the square of the absolute value of 5.3 for all k produces the spectrogram $|X(m,k)|^2$. For conciseness and clarity in the following steps, the notation $|X_m|^2$ is used to represent $|X(m,k)|^2$ of the $m-th$ window.

5.3.3 Spectral Decomposition

The blind signal decomposition of $x[n]$ is then obtained by applying the MSC algorithm to $|X_m|^2$. For each window m, calculate the KDE of $|X_m|^2$, given by

$$p(|X_m|^2) = \frac{1}{N}\sum_{n=1}^{N} K\left(\frac{|X_m|^2 - |X_m|_n^2}{h}\right). \tag{5.4}$$

In Eq. 5.4 $K(u)$ is a kernel function (see section 2.7.2) and $h > 0$ is the kernel bandwidth. The choice of kernel bandwidth has a large influence on the resulting signal decomposition. Further demonstrations will show how h can be estimated as function of the skewness of the resulting noise decomposition spectrum in the absence of prior knowledge. Let $Z(|X_m|^2) = [z_1, z_2, ..., z_I]$ denote the I cluster centers of $|X_m|^2$. Since MSC is non-parametric, the total number of clusters I is not known a-priori. Given $p(|X_m|^2)$, the cluster centers $Z(|X_m|^2)$ can be obtained by applying the iterative mean shift transformation to each point in $|X_m|^2$, given by

$$|X_m|_{\tau+1}^2 = g(|X_m|_\tau^2) \qquad \tau = 0, 1, 2, .. \tag{5.5}$$

where τ denotes the iteration number, and $g(|X_m|^2)$ is the gradient ascent function obtained by setting the gradient of $p(|X_m|^2)$ to zero, resulting in

$$g(|X_m|^2) = \frac{\sum_{n=1}^{N} |X_m|_n^2 K'\left(\left\|\frac{|X_m|^2 - |X_m|_n^2}{h}\right\|^2\right)}{\sum_{n=1}^{N} K'\left(\left\|\frac{|X_m|^2 - |X_m|_n^2}{h}\right\|^2\right)} \tag{5.6}$$

where $K' = dK/du$. Using Eq. 5.6, the mean shift vector $m(|X_m|^2)$ can be defined as

$$m(|X_m|^2) = g(|X_m|^2) - |X_m|^2 \tag{5.7}$$

Using Eq. 5.7 in conjunction with Eq. 5.6, each point in $|X_m|^2$ is shifted towards the local mean until the magnitude of the $m(|X_m|^2)$ is less than the standard error. A more

intuitive way to understand the mean shift algorithm is to revisit the example shown in Fig. 5.1. Recall in the KDE, the regions of high probability density appear as "hills" or peaks, as shown in Fig. 5.1c. Eq. 5.7 and 5.6 simply tells every point in $|X_m|^2$ to climb towards the nearest peak. Over iterations, every point in $|X_m|^2$ will converge into clusters at the dominant peaks in $p(|X_m|^2)$. At convergence, the remaining peaks in $|X_m|^2$ will form the set of cluster centers $Z(|X_m|^2)$, and the peak (i.e. z_i) that each point has to climbed to determines the cluster membership of that point. Spectral decomposition can then be performed by classifying the clusters in $Z(|X_m|^2)$. Applying the same signal assumption in eq. 5.2 to $Z(|X_m|^2)$ produces

$$Z(|X_m|^2) = Z_c(|X_m|^2) + Z_e(|X_m|^2) \tag{5.8}$$

Eq. 5.8 assumes that S-MSC separates the original signal components $c(t)$ and noise processes $e(t)$ into discrete clusters, where Z_c contains all clusters containing harmonic components while Z_e denotes all clusters which contains only noise processes. Classification is performed by sorting Z in order of their population (i.e. number of points belonging to the cluster). As described in section 5.2.2, the implicit biases placed on the composition of machinery vibration spectra can be used to determine the membership of each cluster in Z. More simply put, the typical frequency content in the spectrum can be broadly categorized into two groups: a group consisting of sparse higher energy components or peaks, corresponding to harmonics and other dominant processes in the gearbox, and a second group containing dense lower energy broadband noise and/or non-dominant gearbox processes (Fig. 5.1). Hence, it is assumed that the cluster containing the largest number of points will contain any dense underlying noise processes as well as other low-energy components within the signal, while the remaining clusters will correspond to harmonic components. Formally, this can be expressed by the logic

$$\begin{aligned} Z_i \in Z_c \;\; if \;\; Z_i &\neq \arg\max_{Z_i} \;\; count(Z) \\ Z_i \in Z_e \;\; if \;\; Z_i &= \arg\max_{Z_i} \;\; count(Z) \\ & \qquad\qquad i = 1, 2, ..., I \end{aligned} \tag{5.9}$$

Hence, the points in $|X_m|^2$ belonging to Z_c comprise the S-MSC harmonic spectrum, while the points belonging to Z_e comprise the S-MSC residual spectrum.

5.4 Validation Approach

In order to demonstrate robustness of S-MSC as a pre-processing tool for analysis of non-stationary signals, the validation approach will involve the application of S-MSC to a number of different applications and analysis frameworks, defined as follows:

1. **Extraction of time-varying harmonic content from noise in non-stationary vibration signals:** apply S-MSC to extract time-varying gearbox harmonics in non-stationary LINK APM gearbox vibration data. Comparison of S-MSC to marginal spectrum and TSA [15].

2. **S-MSC as a pre-processing technique for bearing fault detection in non-stationary signals:** apply S-MSC to for de-noising of non-stationary aircraft engine vibration data prior to performing envelope analysis for bearing fault detection in non-stationary signals. This data was obtained from exercise 2 of the Surveillance 8 challenge, which features an accelerating Safran aircraft engine containing two outer race bearing faults [8].

3. **Comparison to other spectral pre-processing approaches:** apply S-MSC to extract time-varying harmonics from a non-stationary wind turbine vibration signal and compare the performance of S-MSC to those obtained using the synchrosqueezing transform [137] and singular spectrum analysis [65]. This data was obtained from the CMMNO 14' challenge [29].

5.4.1 Evaluation Metrics

The appropriate evaluation metric(s) will vary between each validation application. In general, where a-priori knowledge is available (i.e. known gearbox frequencies), the performance of the S-MSC algorithm can be evaluated against/using the known parameters. In addition, while the main objective of the proposed approach is the extraction of time-varying harmonic components in non-stationary vibration signals, the extracted harmonic components are inherently de-noised as a part of the process. For the three former cases, this level of inherent de-noising achieved by the proposed approach is estimated. Often in practice, the signal-to-noise ratio is unknown and varying in the case of non-stationary and time-varying signals. In order to estimate the inherent de-noising performance of the proposed method, all non-harmonic signal components are assumed to be part of the noise process. Hence, at each position of the time window, the energy of the non-harmonic

signal components isolated by MSC is assumed to be equivalent to the local noise level. The global level of noise extracted from the entire signal can then be represented in terms of the average and standard deviation (σ) of the the local noise values obtained at each position of the time-window.

5.5 Application of S-MSC to LNK APM Gearbox Data

5.5.1 Characterization of the LINK APM Gearbox Vibration Data

Examples of the non-stationarity found in the LINK APM gearbox vibration data can be seen from the typical LINK APM gearbox vibration signal and corresponding spectrogram depicted in 5.2 and 5.3, respectively. The signal shown in 5.2 and 5.3 corresponds to one complete round trip cycle of the LINK train. The spectrogram representation of the APM gearbox signal shows that the spectral content of the gearbox is non-stationary and noisy throughout the train's travel. In addition to the amplitude and frequency modulations, the APM signals exhibit many instances where harmonics appear and disappear from the signal entirely. This phenomena appears to be dependent on the position of the train along it's route. This behaviour is illustrated in the zoomed time-frequency representation in Fig. 5.4 and the Fourier transform spectra shown in Fig. 5.5. Fig. 5.5 illustrates the STFT spectra obtained from different time windows of the APM signal, in which several harmonics are present in the signal at one train position, while missing from another. These non-stationarities can be attributed to operational factors, such as the stop and start phases and fluctuating passenger loads; as well as to mechanical factors such as the bi-directionality of the gearbox and variable gradient of the guideway. The results presented hereafter are generated using the vibration data obtained from the input shaft accelerometer (A4 in Fig. 4.7).

S-MSC is applied to the signal shown in Fig. 5.2, using a frequency resolution of 1 Hz. Fig. 5.6 illustrates the original Fourier spectrum and resulting peak (Z_c) and noise (Z_e) spectra for a single STFT window obtained using S-MSC. Fig. 5.6 illustrates the linear filter property of S-MSC, in that the original spectrum is decomposed without any loss of information or non-linear transformation. The complete spectrogram of S-MSC peak component Z_c is shown in Fig. 5.7. The average level of noise removed during the S-MSC process is estimated to be -0.13 dB with $\sigma = 7.5dB$.

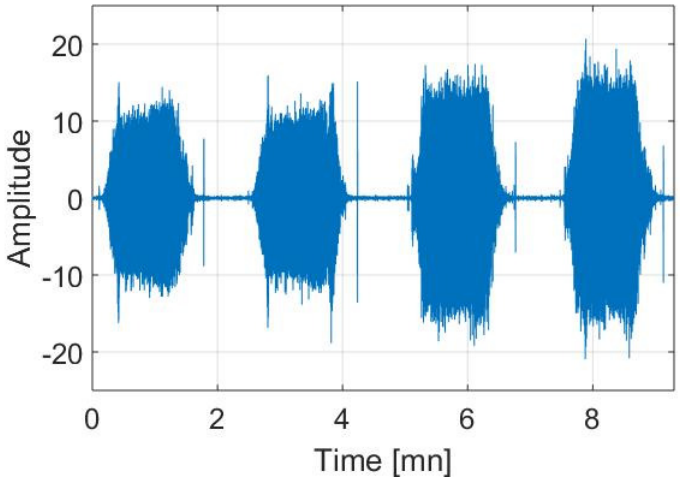

Figure 5.2: Vibration signal for LINK APM gearbox - round trip

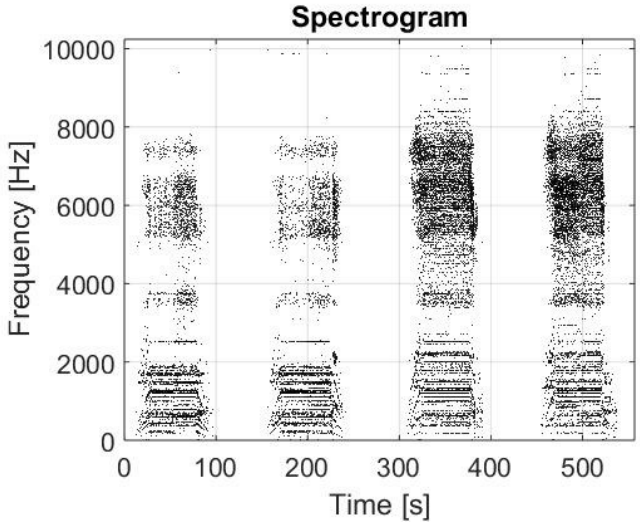

Figure 5.3: Spectrogram of LINK APM gearbox round-trip vibration signal

5.5.2 Comparison to the Marginal Spectrum

A baseline for the performance of S-MSC can be established by comparing the S-MSC Fourier and TF results to those obtained without pre-processing. Fig. 5.8a shows the time-averaged spectrum of the S-MSC peak spectrogram shown in Fig 5.7. For comparison, Fig.

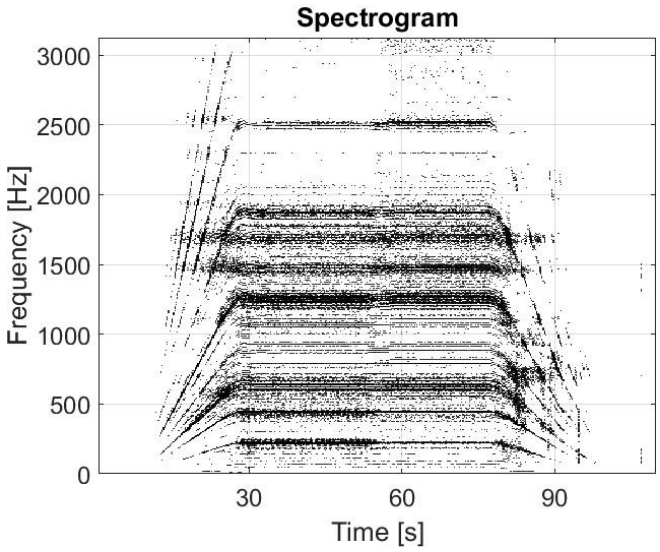

Figure 5.4: Spectrogram for a single stop-start-stop phase of the LINK APM gearbox

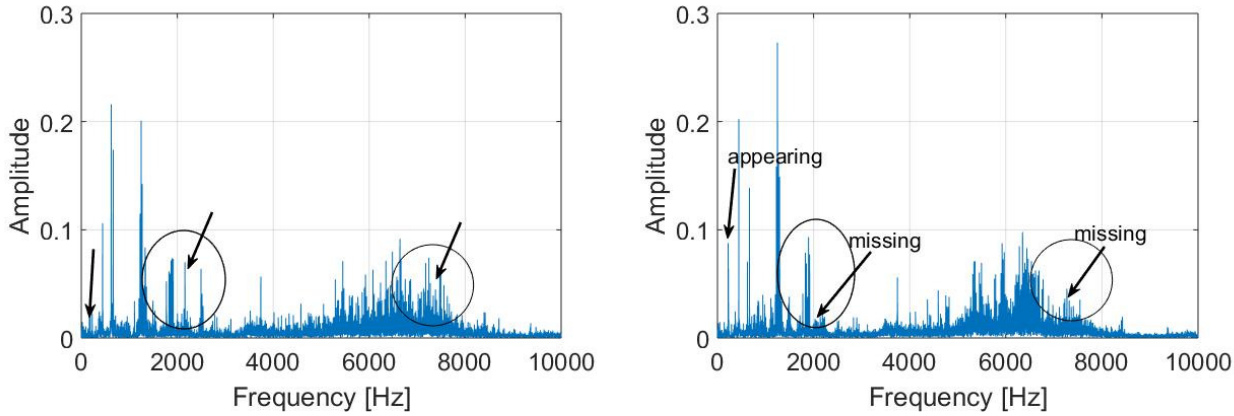

Figure 5.5: Fourier spectra from different segments of the raw LINK APM vibration signal, containing different harmonic components

5.8b illustrates the marginal spectrum distribution of the entire vibration signal, which is obtained by time-averaging the spectrogram shown in Fig. 5.3 without pre-processing using S-MSC. The following analysis references the LINK APM gearbox frequencies contained in Table 4.1 in Section 4.4.1. Direct comparison between the two spectra shows that the spectrum obtained using S-MSC contains more harmonic spectral content than its marginal counterpart. While both spectra contain spectral components related to the gear meshing

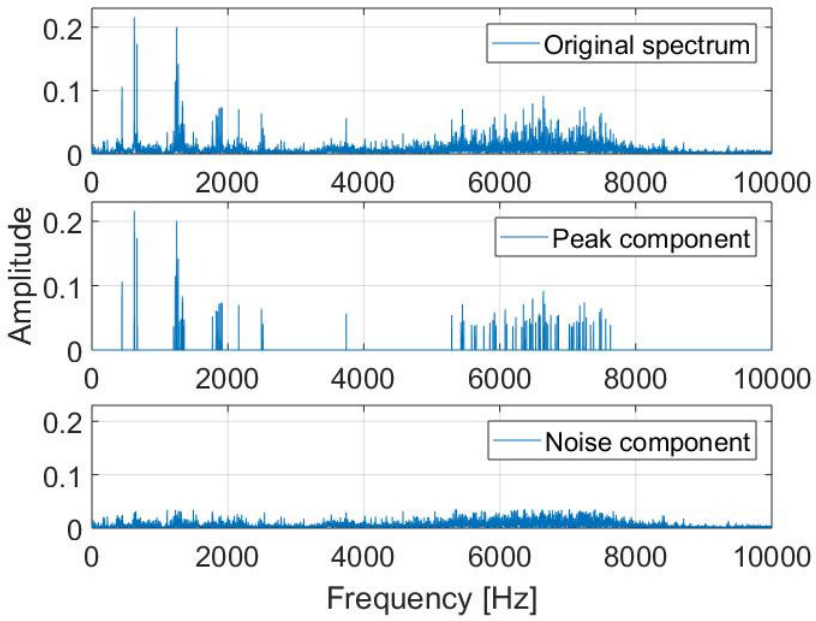

Figure 5.6: Typical decomposition of a frequency spectrum into peak and noise components using Mean Shift Clustering

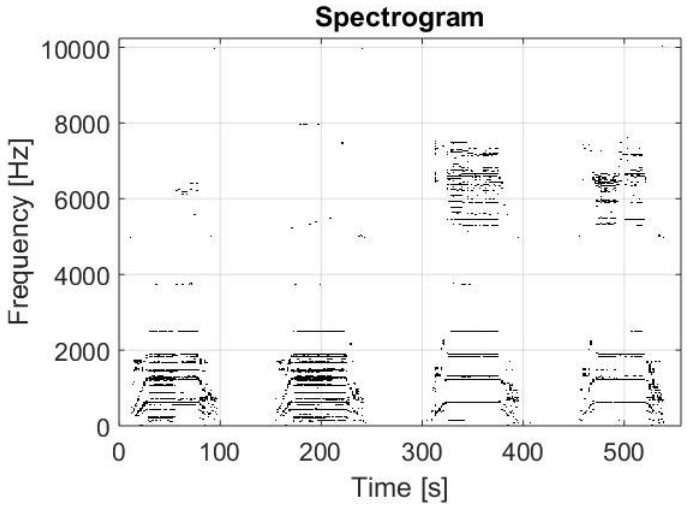

Figure 5.7: De-noised spectrogram of LINK APM gearbox round-trip signal after mean shift clustering

families from gearbox stage a and b, the S-MSC spectrum was able to capture additional harmonics from these families, as well as a gear mesh harmonic from gearbox stage c that is absent in the marginal spectrum. Specifically, the enhanced spectrum in Fig. 5.8a captures the harmonics $h_c^3, h_b^3, h_b^4, h_a^6$, which are missing from the marginal counterpart. Furthermore, the resonant region around 6-8 kHz is captured in the S-MSC spectra, and far less visible in the marginal spectra.

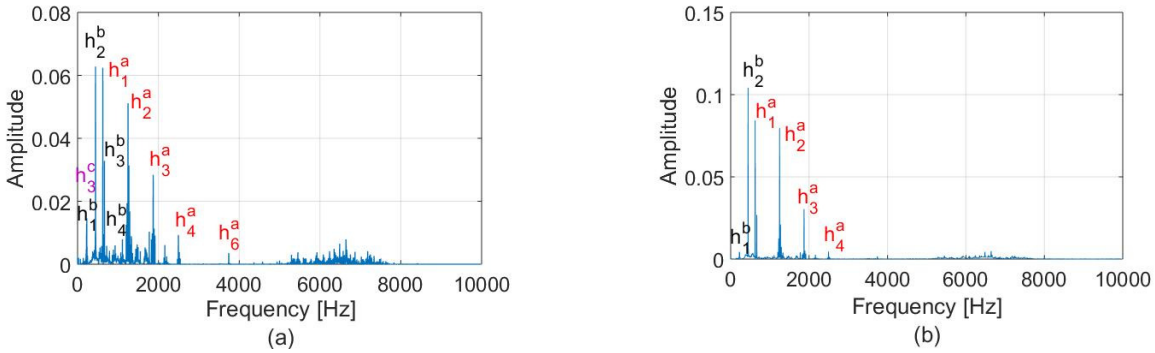

Figure 5.8: (a) S-MSC spectrum obtained using the proposed approach in conjunction with time averaging (b) Marginal frequency spectrum obtained from the signal spectrogram - the extracted gearmesh components labeled

Further comparisons between the S-MSC and marginal spectra can be made from zoomed-in depictions shown in Fig. 5.9a and b. Comparison of these spectra again shows that the S-MSC spectrum contains frequency information not present in the marginal counterpart, including frequency components related to the shaft rotational frequency. Other prominent spectral structures, such as modulation sidebands, are also captured more clearly by S-MSC. Fig. 5.10a shows that the modulation sidebands surrounding the second harmonic from gear mesh stage B were also captured in the S-MSC spectrum, but missing from the mariginal spectrum shown in Fig. 5.10b.

5.5.3 Comparison to Time Synchronous Averaging

Order-tracked vibration data was sampled from the LINK APM gearbox to obtain the TSA signal. The corresponding TSA spectrum for the APM signal was computed using the TSA algorithm presented in [15], which applies re-sampling of the time domain data in order to account for non-stationary shaft speed. The TSA spectrum corresponding to the input shaft, is shown in Fig. 5.11, for data sampled at 200 kHz and 20 kHz.

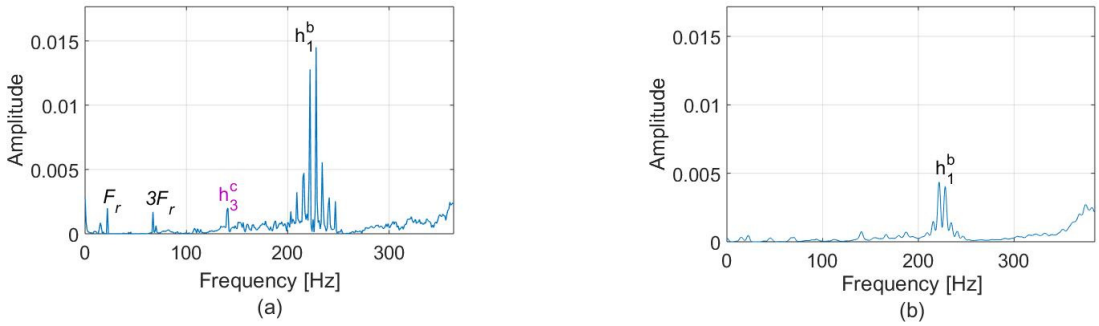

Figure 5.9: Comparison in the 0-400 Hz region: rotational frequency, gearmesh harmonics and modulation sidebands extracted using (a) S-MSC, (b) the marginal frequency spectrum

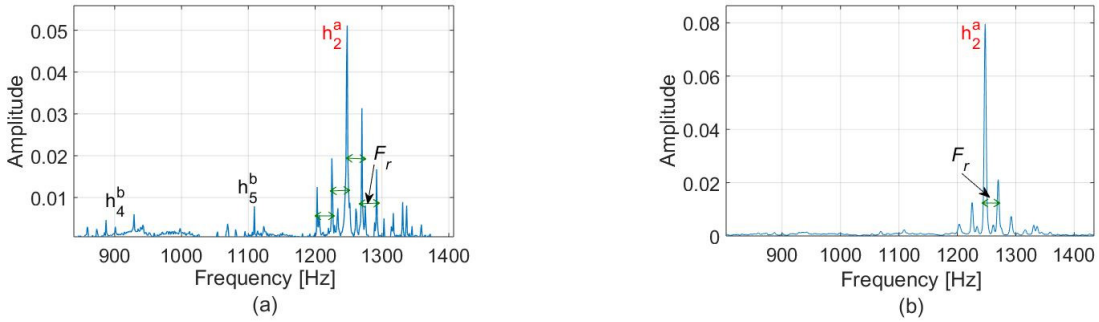

Figure 5.10: Comparison in the 800-100 Hz region: gearmesh harmonics and modulation sidebands extracted using (a) S-MSC (b) the marginal frequency spectrum

Comparison of the 200 kHz SA spectrum with the S-MSC spectrum in Fig 5.8a shows that both spectra contain the same harmonic content corresponding to the input shaft, with very similar noise levels. However, the S-MSC has several advantages over TSA for the type of vibration data considered. First, when using TSA, any frequency components that are asynchronous to the selected fundamental frequency (usually shaft speed) are removed in the process. Hence, the spectrum corresponding to each gearbox stage must be obtained separately, as opposed to the proposed spectral analysis technique, where all of the spectral information is obtained in single step. Second, TSA needs an additional angular reference signal and requies the kinematic information (i.e. gear ratios) for the gearbox must be known a-priori for TSA. Third, comparison of the spectra shown in Fig. 5.11a and b illustrates the influence of sampling rate on the frequency resolution when using SA. Closely spaced harmonics and sidebands, which are visible in the S-MSC spectrum shown in Fig. 5.8a, are not visible in the TSA spectrum for data sampled at the same

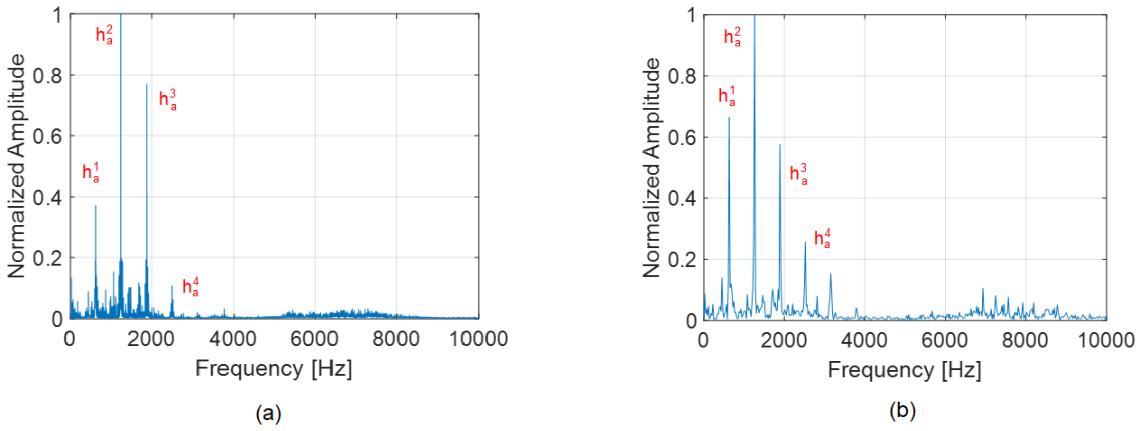

Figure 5.11: Frequency spectrum obtained using time synchronous averaging where the sampling frequency is (a) 200 kHz, (b) 20 kHz

rate. Comparable frequency resolution between the two methods could only be obtained for TSA data sampled at 200kHz, which would increases the data collection requirements for TSA by an order of magnitude relative to S-MSC. Finally, from a practicality angle, the presence of the significant speed fluctuations within the APM gearbox presented a challenge for obtaining suitable data for TSA. Without the use of more sophisticated pre-processing techniques prior to TSA, the TSA spectrum can only be obtained in regions where the speed fluctuations are nominal, whereas S-MSC imposes no restriction on the type or level of speed fluctuations in the signal.

5.6 Application of S-MSC to Safran Aircraft Engine Data

The main challenge posed to bearing fault detection by the Safran aircraft engine data is that the constantly changed speed results in bearing fault frequencies that are function of time, rather than constant. A common diagnostic approach for this type of data is to first obtain an estimate of the shaft speed, which can be then be used in a subsequent pre-processing step, such as angular resampling, to obtain a stationary representation of the signal which can be used for bearing fault analysis. A key observation from the Surveillance 8 challenge debriefing [8] is that all of the successful approaches relied upon traditional, computationally expensive and parametric pre-processing methods to perform this

speed estimation. This section is intended to demonstrate how S-MSC can be used as an efficient, non-parametric substitute to the computationally expensive and parametric pre-processing approaches found in most existing fault diagnostic approaches for non-stationary signals. Specifically, this section illustrates the results obtained when using S-MSC as a pre-processing tool in combination with envelope analysis 2.5.2 for detection of bearing faults in the Surveillance 8 Safran Aircraft engine data.

5.6.1 Dataset Description

The data used was obtained from exercise 2 of the Surveillance 8 challenge. A full description of the engine, gearbox kinematics and data acquisition process can be found in [8]. The vibration data was sampled from an accelerating Safran aircraft engine with a sample rate of 50 kHz over a duration of 200s. The engine contained two faulty bearings (L1 and L5) with outer race faults. Fig. 5.12 illustrates the speed of the shaft containing the damaged bearing (L5), and shows non-stationarity due to the acceleration of the engine over the sampling interval.

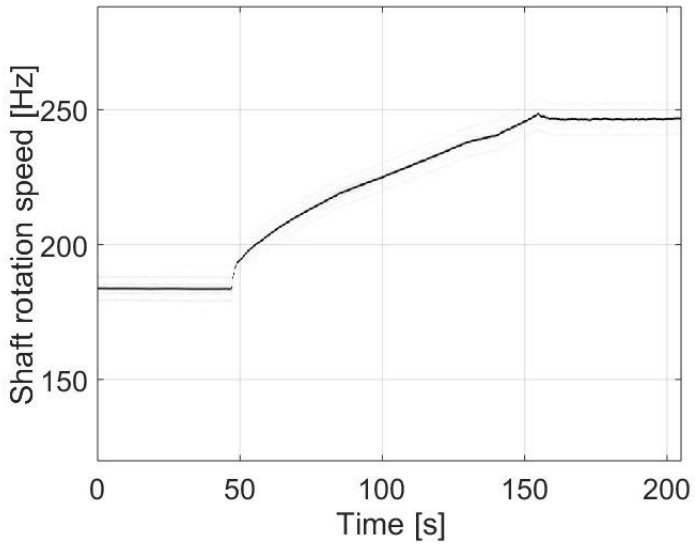

Figure 5.12: Rotational speed of shaft containing damaged bearing L5

5.6.2 Procedure

The S-MSC bearing fault detection procedure is as follows:

1. Segment the signal into 2s windows and calculate the STFT and corresponding PSD of each window according to Eq. 5.3.
2. For each window:
   (a) Compute the KDE of the PSD (Eq. 5.4)
   (b) Apply the mean shift clustering to Eq. 5.4 using Eqs. 5.5, 5.6 and 5.7 and decompose the spectrum using Eq. 5.9 to obtain Z_c and Z_e
   (c) Band-pass filter Z_c at the resonant frequency band of 21.5-24.8 kHz
   (d) Demodulate the band-pass filtered signal Z_c and obtain the envelope signal $|\bar{x}(t)|$ using the HT according to Eq. 2.22
   (e) Compute the envelope spectrum of $|\bar{x}(t)|$
3. Combine the envelope spectra of $|\bar{x}(t)|$ into a single spectrogram representation.

5.6.3 Envelope Analysis Results

The performance of the S-MSC was assessed by comparing the envelope spectrograms obtained with and without the application of S-MSC. Fig. 5.13a illustrates the envelope spectrogram without S-MSC, while Fig. 5.13b illustrates the spectrogram produced using the procedure outlined in section 5.6.2. Comparison of the two spectrograms shows that the raw envelope spectrogram contains a high level of background noise, resulting in only faint traces of the outer race bearing fault frequency (BPFO) and its second harmonic (BPFO 1 and 2) appearing in the spectrogram. However, the de-noised S-MSC envelope spectrogram shows relatively sharp traces of the outer race fault frequency along with two harmonics (BPFO 1, 2 and 3) as well as the corresponding modulation sidebands - which can be used to diagnosis the bearing fault. Furthermore, the cage frequency (CF), and its corresponding sidebands are also visible in the S-MSC spectrogram. Comparing the noise of the two resulting spectrograms, it was found that the S-MSC spectrogram contained 18 dB less noise than the raw counterpart, indicating a significant improvement to the SNR after the application of the proposed approach. These results demonstrate that S-MSC can be used as an efficient, non-parametric alternative to traditional pre-processing tools for bearing fault detection in non-stationary signals.

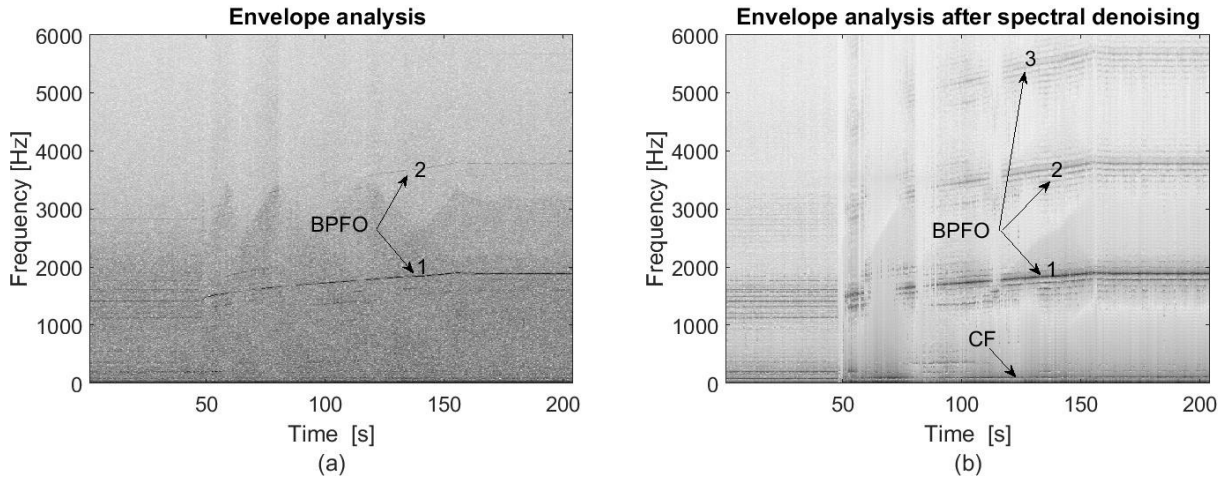

Figure 5.13: Envelope spectrogram of vibration signal: (a) without and (b) with spectral de-noising using Mean Shift Clustering

5.7 Application of S-MSC to CMMNO'14 Wind Turbine Data

The purpose of this section is to compare the performance of S-MSC to two state-of-the-art spectral pre-processing approaches: the Fourier Synchro Squeezing Transform (FSST) [137] and sliding-Singular Spectrum Analysis (SSA) [65]. The analysis is conducted using data obtained from the CMMNO' 14 wind turbine dataset consists of non-stationary wind turbine gearbox data. FSST is a spectral pre-processing tool for non-stationary signals that can be used to express an input signal as a sum of analytic signals [137]. The principle of FSST is similar to that of frequency reassignment, in that FSST localizes energy surrounding a harmonic to the harmonic, which produces an enhanced TF representation. Like S-MSC, FSST operates on the spectrogram of the signal obtained using the STFT. Unlike S-MSC however, FSST is not conducive to online CM applications due to its computationally heavy process. FSST has been used applied to the analysis of non-stationary wind turbine signals and hence, is included here for the comparison against S-MSC. Sliding SSA is a parametric spectral pre-processing tool that utilizes singular value decomposition (SVD) in conjunction with a sliding window to decompose a non-stationary signal into physically interpretable components representing oscillations and noise [65]. According to SSA-based de-noising, the separation of signal harmonics from noise is achieved by selecting the first n pairs of singular values, which are sorted in decreasing order, and reconstructing the signal using the associated singular vectors, where n is the number of harmonics

perceived to be in the signal.

5.7.1 Dataset Description

The CMMNO' 14 wind turbine dataset consists of non-stationary wind turbine gearbox data sampled at 20kHz. The wind turbine contains a inner ring fault on the planet bearing. A detailed description of the gearbox and test parameters can be found in [99]. The wind turbine signal to be analyzed consists of multiple families of deterministic frequencies embedded within broadband noise, which are subjected to frequent non-linear speed fluctuations. Wind turbine signals contain multiple sources of stationarity stemming from time-dependent, angle-dependent and time and angle-independent phenonena [99].

5.7.2 Comparison to sliding SSA

The raw time-frequency spectrogram of the wind turbine signal is shown in Fig. 5.14a, which was generated using overlapping 2s windows (75% overlap) and a frequency resolution of 1Hz. Fig.5.14b shows the S-MSC de-noised spectrogram generated using the same parameters, which shows that the time-varying harmonic components were successfully separated from the broadband noise, despite the presence of heavy non-linear speed fluctuations. The average level of noise removed from the signal using S-MSC is estimated to be 1.45 dB, with $\sigma = -11dB$.

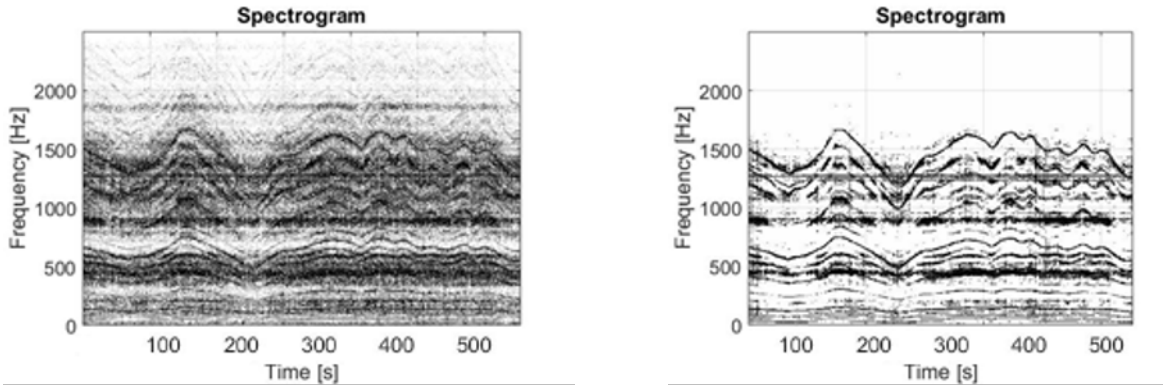

Figure 5.14: Spectrogram of wind turbine vibration signal: (a) without and (b) with spectral de-noising using Mean Shift Clustering

Since SSA is a parametric approach, the spectral information shown in Fig. 5.14a was used to determine the optimal value of $n = 20$. Applying sliding SSA to the wind turbine signal using the same window parameters produces the de-noised spectrogram shown in Fig. 5.15. The level of noise removed using the SSA approach is estimated to be -18 dB, with standard deviation $\sigma = 1dB$. Comparison of the SSA spectrogram shown in Fig. 5.15 to S-MSC spectrogram in Fig. 5.14b shows both methods are equally effective at highlighting the signal harmonics buried in noise. However, the S-MSC spectrogram contains less residual noise and sharper harmonics than the SSA counterpart. Additionally, unlike MSC, sliding SSA involves the use of the computationally expensive SVD algorithm, which makes sliding SSA less conducive for online CBM applications.

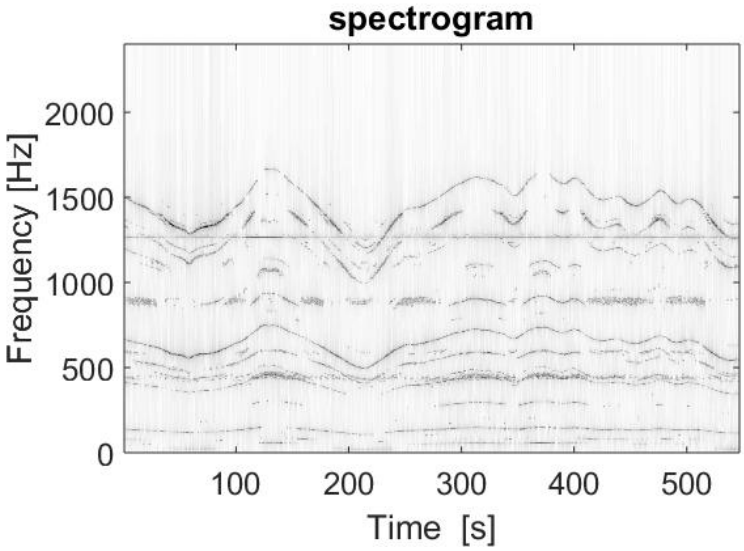

Figure 5.15: Spectrogram of wind turbine vibration signal with spectral de-noising using sliding SSA

5.7.3 Comparison to FSST

Based on the assumed nature of the frequency content in the signal, FSST was applied to the wind turbine signal using both a Kaiser and Hanning window, and the optimal window type and window length N was determined using an iterative approach. In all cases, a window overlap of $N - 1$ was used. For larger values of N, the signal was batch processed due to memory constraints. The best TF representation was obtained using the FSST with

a Kaiser window with $\beta = 10$ and $N = 512$. From Fig. 5.16, it can be seen that FSST struggles to localize the energy around the harmonics due to the low SNR of the signal. While the wavelet based SST and generalized SST approach in [1] may perform better on this type of signal, they are not conducive to online monitoring, which is one of the primary motivations behind the proposed MSC approach. These results illustrate the robustness of the proposed approach, and suggests that the MSC approach is insensitive to the type of speed fluctuations commonly found in non-stationary systems such as wind turbines. For these types of signals, the performance of the proposed approach in both the extraction of time-varying harmonic components and de-noising was found to be comparable to that of Sliding SSA, while requiring significantly less computational resources.

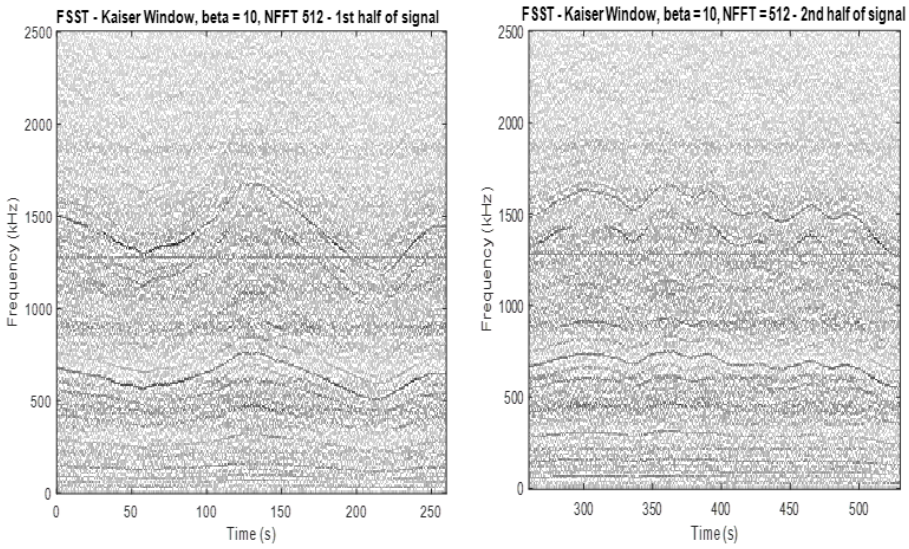

Figure 5.16: Spectrogram of wind turbine vibration signal with spectral pre-proccesing using FSST

5.8 Evaluation of the Kernel Bandwidth Parameter

The choice of KDE kernel bandwidth parameter h controls the behaviour of the S-MSC spectral decomposition. The value of h directly influences the smoothness of KDE surface over the data set, which can affect the clustering result obtained using S-MSC [32, 25]. The topic of bandwidth estimation has been extensively studied, resulting in a number of

data-driven estimators associated with different optimization criteria which can be used to estimate h [32]. If specific prior information is known, such as the true probability distribution of the data, h can simply be estimated by minimizing the RMSE. Conversely, if a-priori information is not available, numerical approaches such as cross-validation, or application-specific criterion can be used to select the kernel bandwidth.

In the present context, S-MSC is used as a pre-processing approach to separate time-varying harmonic components from noise in non-stationary machinery vibration signals. In the absence of prior knowledge, there currently does not exist a suitable criterion to obtain a blind estimate h for spectral decomposition of machinery vibration spectra. Hence, this section presents the results of a heuristic study which was used to determine a rule-of-thumb estimator for h in the context of machinery vibration signals. The proposed estimator, which is based on the spectral skewness of the decomposition component Z_e, can be used to determine a suitable value for h in the absence of prior knowledge. The Surveillance 8 Safran and CMMNO'14 wind turbine vibration signals were considered for the study.

5.8.1 Results of Bandwidth Study

For each signal, an arbitrary 3s window of the data was selected, and the S-MSC is applied to the each window across a broad range of bandwidth values. For each value of h, the power and spectral skewness of the resulting decomposition spectra Z_c and Z_e is calculated. Spectral skewness is analogous to skewness for a time signal, and is defined as the third central moment of the Fourier spectrum, which can be expressed as

$$\begin{aligned} \textit{spectral skewness} = E[(|X_m|^2 - E[|X_m|^2)^3] &= \frac{\sum_{k=1}^{K}(k-\mu_1)^3|X_{m_k}|^2}{\mu_2^3\sum_{k=1}^{K}|X_{m_k}|^2} \\ \mu_1 &= \frac{\sum_{k=1}^{K}k|X_{m_k}|^2}{\sum_{k=1}^{K}|X_{m_k}|^2} \\ \mu_2 &= \sqrt{\frac{\sum_{k=1}^{K}(k-\mu_1)^2|X_{m_k}|}{\sum_{k=1}^{K}|X_{m_k}|^2}} \end{aligned} \tag{5.10}$$

For each signal, the resulting spectral skewness and power of each window are shown in Fig. 5.17a and b, respectively. An important observation from Fig. 5.17 is that the skewness of a spectrum consisting of purely harmonic components is much higher than

the skewness of the spectrum consisting of pure noise. Using this principle, the following hypothesis is proposed: the value of h which minimizes the spectral skewness of the residual spectrum Z_e produces the optimal decomposition result. Based on Figs. 5.17a and b, there appears to be an optimal region for the h for both signals in which the skewness of the Z_e is minimized.

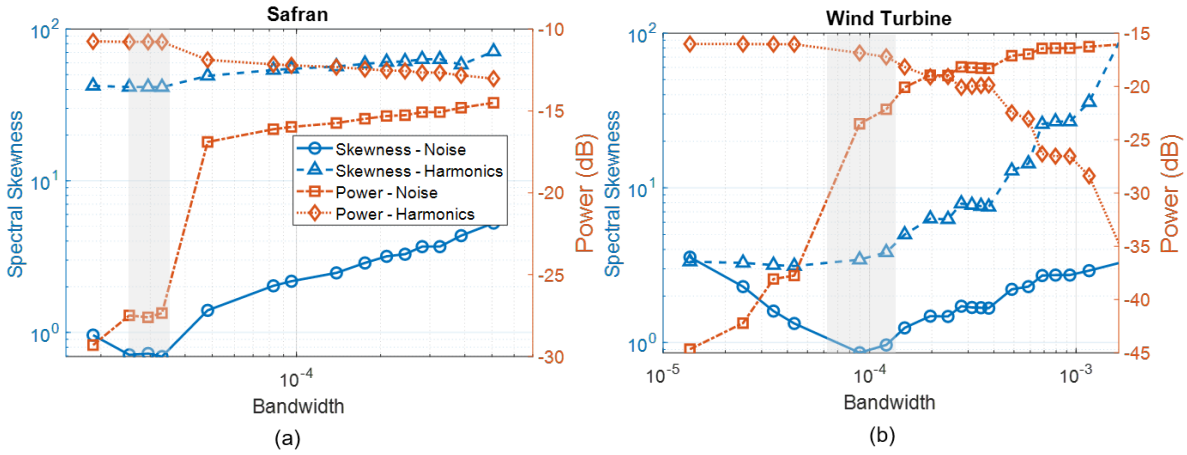

Figure 5.17: Effect of bandwidth on skewness and power of MSC decomposition spectra (a) Safran (b) Wind Turbine

Fig. 5.18 shows the decomposition spectra for the wind turbine signal using the bandwidth corresponding to the minimum skewness, illustrating an effective separation of harmonic and noise components. Fig. 5.19a-b shows the decomposition spectra corresponding to the smallest and largest bandwidth values for the wind turbine data, respectively. In the case of the smallest bandwidth, the over-decomposition of the noise component is apparent, leading to the minimum noise power level shown in Fig. 5.19a. In the case of the largest bandwidth, the decomposition is unable to separate harmonic components from noise, resulting in the residual spectrum shown in Fig. 5.19b, which contains both harmonic and noise components. The over-decomposition of noise can also be seen in power trend of the decomposition spectra, as shown in Fig 5.17. When considering the range of bandwidths to the left of the optimum region, the power of both decomposition spectra can be seen to change rapidly with respect to the bandwidth, indicating a significant amount of frequency components (presumably noise) are transferred between the Z_c and Z_e as the bandwidth increases. This behaviour continues until the bandwidth approaches the lower bound of the optimal region, at which point the skewness of the residual spectrum approaches a minimum, at which point it is presumed that the majority of the noise in the signal has

been separated from the harmonic component.

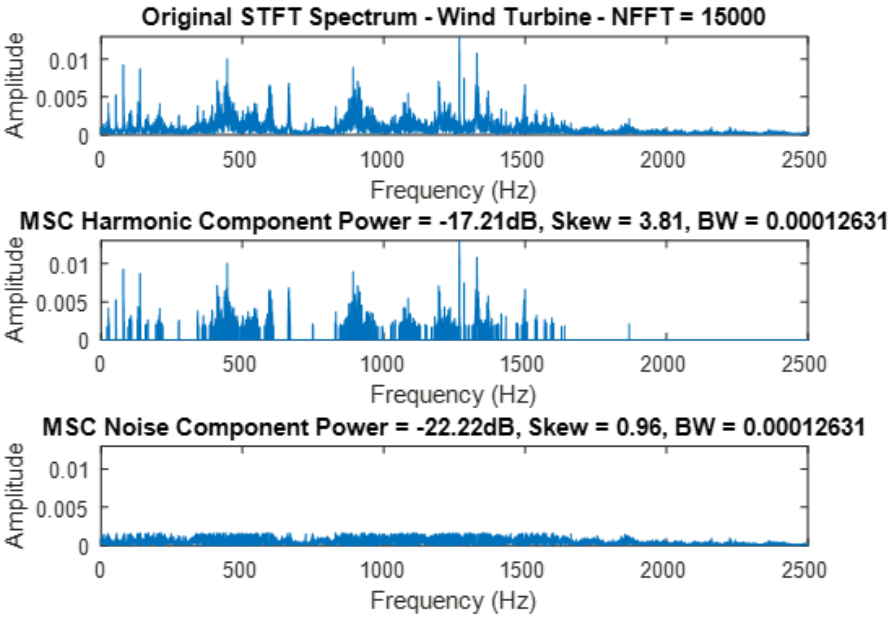

Figure 5.18: MSC decomposition of wind turbine signal corresponding to optimal bandwidth

The results of the bandwidth study demonstrate that for machinery vibration spectra, smaller values of h increase the sensitivity of the decomposition with respect to $|X_m|^2$, which allows for a larger number of clusters and separation of weaker harmonic components situated closer to the noise floor. However, a lower bound exists in which the bandwidth will result in over-decomposition of the noise. On the contrary, larger bandwidths can only capture significant differences in frequency amplitudes, resulting in clusters which contain mixtures of harmonic and noise components. In both cases, selecting the kernel bandwidth h based on the minimization of the spectral skewness of Z_e provides a reasonable rule-of-thumb estimator which can be used in the absence of prior information. Formally, this can be expressed as

$$h_{opt} = \arg\min_{h} \quad E[(|X_m|^2 - E[|X_m|^2)^3] \tag{5.11}$$

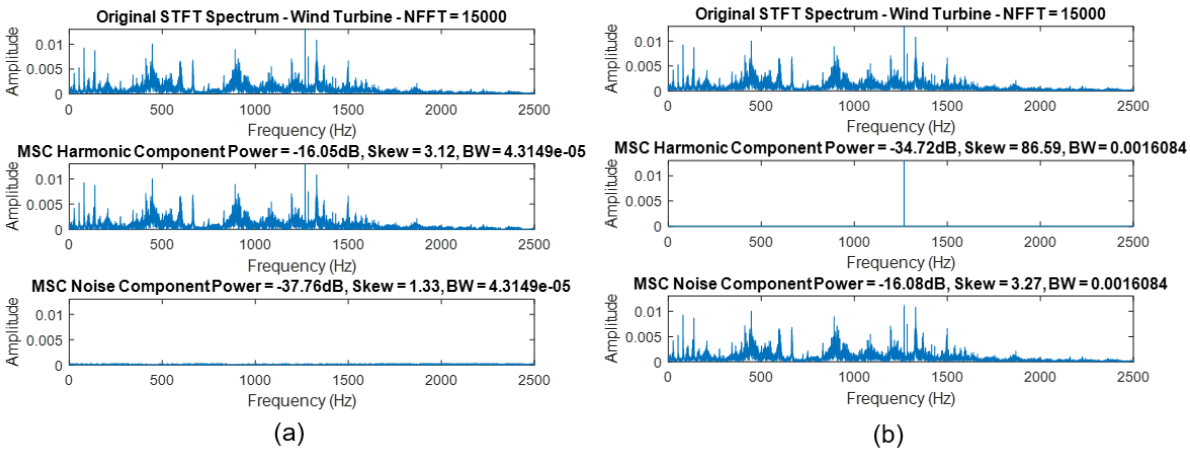

Figure 5.19: S-MSC decomposition of wind turbine signal using extreme values of h (a) lower bound (b) upper bound

5.9 Summary

This chapter presents a model-free, blind spectral analysis approach for pre-processing non-stationary vibration signals. The approach consists of applying MSC to the STFT of vibration signals in order to simultaneously de-noise and separate time-varying harmonics from noise. Validation was performed across multiple use cases and data sets. The key conclusions from this work are as follows:

1. Validation performed on the LINK APM gearbox data showed that S-MSC was effective at separating the time-varying gearbox harmonics from noise. Comparison to pre-processing using TSA showed that S-MSC was able to extract more harmonic components than TSA, without requiring the angular signal needed for TSA.

2. Application to the Safran engine data set showed that the use of S-MSC for de-noising in envelope analysis enabled the detection of time-varying bearing fault frequencies, circumventing the need for alternative, computationally expensive solutions.

3. Further validation against FSST and SSA using the CMMNO'14 wind turbine data showed that S-MSC performs better than FSST and equal to SSA on the low-SNR CMMNO'14 data, at a much lower computational cost.

4. The S-MSC approach depends upon only a single parameter, the kernel bandwidth.

To enable autonomous deployment of S-MSC, a criterion based on the spectral skewness of the S-MSC residual spectrum was proposed for the estimation of the kernel bandwidth.

Chapter 6

Gaussian Mixture Model Operating State Decomposition (GMM-OSD)

This chapter presents a soft-clustering pre-processing tool for blind separation of sensor data obtained from machinery over multiple operating states, termed Gaussian Mixture Model Operating State Decomposition (GMM-OSD). The purpose of GMM-OSD is to provide an unsupervised means to decompose a non-stationary multi-state vibration signal—for which many conventional vibration analysis and fault detection techniques are not suitable—into a set of stationary, constituent signals to which existing tools can be readily applied. In GMM-OSD, each of the resulting decomposition signals corresponds to a discrete operating state. Using the fitted model, the unknown operating state of a new observation can be classified using posterior inference. GMM-OSD can also be used in conjunction with other unsupervised pre-processing tools (i.e. S-MSC) to form a robust pre-processing framework suitable for the complex, dynamic operational behaviour of many modern engineering systems. Hence, the method presented in this chapter addresses the key gap in CBM literature pertaining to the lack of computationally efficient, unsupervised pre-processing algorithms to perform operating state-decomposition.

This chapter begins by presenting the motivations for GMM-OSD, followed by Section 6.1, which presents the required background and methodology. Section 6.3 and onwards presents the results from validation of the proposed approach performed on experimental data collected from the Pearson LINK APM and passenger boarding tunnel PCA field pilots.

6.1 Motivations

As shown in Section 4.1, many types of rotating machinery in industrial applications, such as APMs and pre-conditioned air units, have more than one operational state. These changes in operational behaviour are reflected in their vibration behaviour as well, which results in a non-stationary signal that is not conducive to many common vibration analysis and fault detection tools, such as statistical process control (SPC). Applying these tools to multi-state systems involves either first decomposing the signal into a set of stationary, constituent signals to which these tools can be easily applied, or, applying real-time knowledge of the machine operating status to improve the accuracy of the analysis or fault detection results. When prior information, such as machine status is made available to the CBM framework, implementing either of these approaches becomes a straightforward process. However, in the case where such data is not available, and information pertaining to the nature or exact number of the operating states is not known a-priori, this classification task becomes non-trivial. This problem is further complicated for systems containing additional forms of non-stationarity as well (ex: Section 4.1.3), which typically requires additional pre-processing to resolve. GMM-OSD addresses this challenge by presenting a robust means for unsupervised modeling and classification of multi-state machinery vibration data, which can be implemented alongside other pre-processing tools (i.e. S-MSC) for unsupervised condition monitoring of modern, dynamic engineering systems.

The GMM-OSD approach employs GMMs in conjunction with CIs extracted from the vibration signal to classify each data point in a signal collected over multiple machinery operating states into its respective operating state. In the CBM analytical framework, GMM-OSD is applied in the pre-processing step after feature extraction preceding model training, and can be used in conjunction with other pre-processing tools, such as S-MSC. Fig. 6.1 provides an example of a CBM signal processing chain incorporating both GMM-OSD and S-MSC. Model selection without the use of prior knowledge is demonstrated using two approaches–an information criterion-based approach, and a variational Bayesian approach. The ability to automatically determine the optimal number of classes (i.e. operating states) without parameter tuning or user intervention makes GMM-OSD easily deployable in large-scale CBM applications with limited a priori information. In addition, the underlying Gaussian assumption in GMM-OSD makes the models easily interpretable and conducive to subsequent analyses that benefit from knowing the distribution of the data, e.g., Gaussian process regression or hierarchical Bayesian updating.

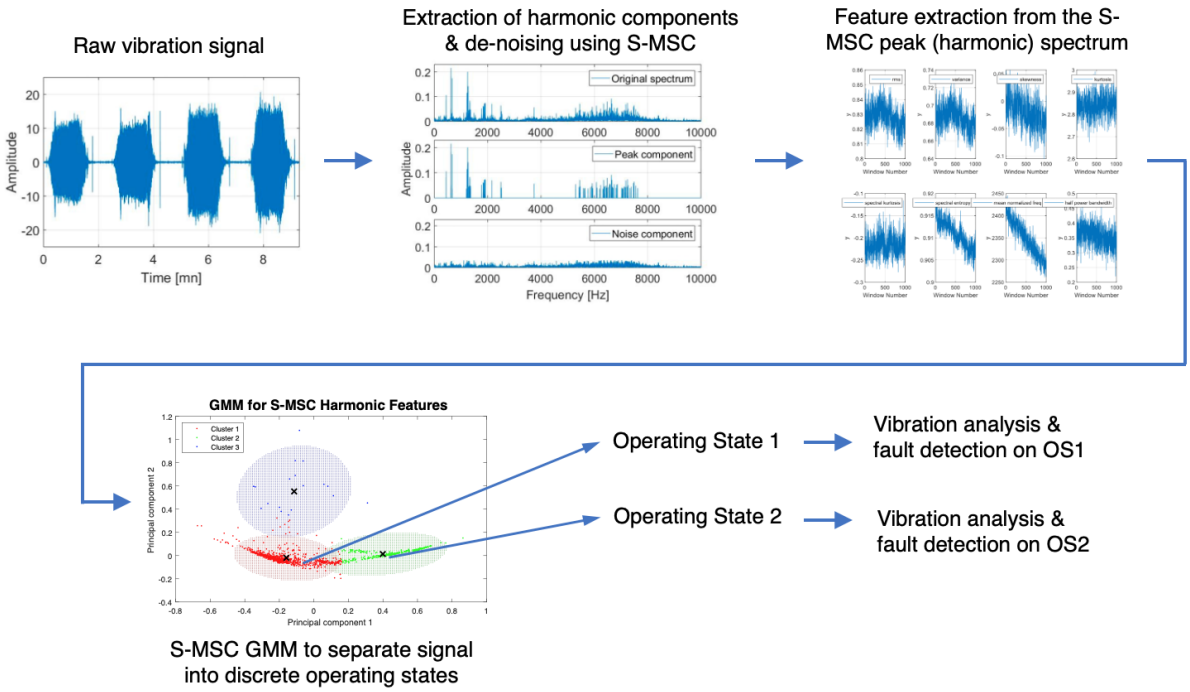

Figure 6.1: Example CBM signal processing chain using GMM-OSD in conjunction with S-MSC pre-processing

6.2 Background and Methodology

This section presents the background and methodology pertaining to GMM-OSD.

6.2.1 Gaussian Mixture Models

Gaussian Mixture Modeling is soft-clustering approach that assumes that a set of data was generated using a group (or mixture) of Gaussian processes. Unlike its hard-clustering K-means counterpart, which assigns a definitive label to each data point, GMM performs soft clustering on the data, which instead assigns a probability of cluster membership of each point instead. The latter approach is more conservative since it captures the uncertainty of the cluster assignment, and hence, is better suited to the unsupervised setting where the ground truth labels of the data points are not known. The central

assumption in GMMs is that the underlying data is generated by some unknown mixture of K Gaussian distributions. Each individual Gaussian distribution k is described by the Gaussian parameters μ_k and Σ_k, as well as a mixture weight π_k, which is used to individual Gaussian distrubtions in the mixture to one another. π_k has the following property

$$\sum_{k=1}^{K} \pi_k = 1 \tag{6.1}$$

Given a data set $X = \{x_1, x_2, ..., x_N\} \in \mathbb{R}^d$, the probability that a data point x_i belongs to a Gaussian distribution k can be expressed as

$$p(z_{ik} = 1|x_i) \tag{6.2}$$

where z is a latent variable belonging to the set $Z = z_1, z_2, ..., z_K$. z_{ik} is equal to 1 if x_i belongs to Gaussian k, and zero otherwise. Using this property, following relationship between z and the Gaussian mixture weights π_k is defined:

$$\pi_k = p(z_k = 1) \tag{6.3}$$

Then, the marginal probability of a point x_i for the multivariate case can be expressed as

$$\begin{aligned} p(x_i) &= \sum_{k=1}^{K} p(x_i|z)p(z) = \sum_{k=1}^{K} \mathcal{N}(x_i|\mu_k, \Sigma_k)\pi_k \\ N(x_i|\mu_k, \Sigma_k) &= \frac{1}{\sqrt{|\Sigma_k|(2\varpi)^K}} exp\left(-\frac{(x_i - \mu_i)^T}{\Sigma_k(x_i - \mu_k)}\right) \end{aligned} \tag{6.4}$$

Using Eq. 6.4, the marginal probability $p(X)$ is given as

$$p(X) = \prod_{i=1}^{N} p(x_i) = \prod_{i=1}^{N} \sum_{k=1}^{K} \mathcal{N}(x_i|\mu_k, \Sigma_k)\pi_k \tag{6.5}$$

Let $\theta = \{\pi, \mu, \sigma\}$ represent the unknown GMM model parameters. Then, given K, the optimal model values of θ are those which maximize the log-likelihood of $p(X)$, given by

$$L(X|\theta) = log(p(X)) = \sum_{i=1}^{N} log\left(\sum_{k=1}^{K} \mathcal{N}(x_i|\mu_k, \Sigma_k)\pi_k\right) \tag{6.6}$$

6.2.2 Expectation Maximization Algorithm

Eq. 6.6 has no closed-form analytical solution, and is typically evaluated numerically using the expectation maximization (EM) algorithm. The EM algorithm is an iterative, two step structure to solve for the maximum likelihood. Let θ^t represent the current estimate of θ, and θ^{t-1} represent the subsequent estimate of θ obtained from the previous EM iteration. At each iteration, the expectation step computes the expected value of the log-likelihood of θ^t over Z given X (i.e. posterior probability of the data being generated by the current Gaussian mixture), given by

$$\mathcal{Q}(\theta^t, \theta^{t-1}) = \mathbb{E}[log(p(X, Z|\theta^t)] = \sum_Z p(Z|X, \theta^{t-1}) log(p(X, Z|\theta^t)) \tag{6.7}$$

To evaluate an expression for $p(Z|X, \theta^{t-1})$ can be derived as a function of z_{ik} using Bayes rule

$$\begin{aligned} p(Z|X, \theta^{t-1}) &= \frac{p(x_i|z_k = 1)p(z_k = 1)}{\sum_{j=1}^K p(x_i|z_j = 1)p(z_j = 1)} \\ &= \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)} = \gamma(z_{ik}) \end{aligned} \tag{6.8}$$

as well as an expression for $p(X, Z|\theta^t)$, which is the likelihood with respect to both X and Z, given in log form as

$$log(p\left(X, Z|\theta^t\right) = \sum_{i=1}^N \sum_{k=1}^K z_{ik} \left[log\pi_k + log\mathcal{N}\left(x_i \mid \mu_k, \Sigma_k\right)\right] \tag{6.9}$$

Substituting Eq. 6.8 and 6.9 into Eq. 6.2.2 results in

$$\mathcal{Q}(\theta^t, \theta^{t-1}) = \sum_{i=1}^N \sum_{k=1}^K \gamma(z_{ik}) \left[log\pi_k + log\mathcal{N}\left(x_i \mid \mu_k, \Sigma_k\right)\right] \tag{6.10}$$

Using Eq. 6.10 calculated in the expectation step, the maximization step solves for the updated model parameters θ^t which maximize the the Eq. 6.10. To simplify the optimization, the previously-defined property of π (i.e. $\sum_{k=1}^K \pi_k = 1$) is applied as a constraint to Eq. 6.10 which yields

$$\mathcal{Q}(\theta^t, \theta^{t-1}) = \sum_{i=1}^N \sum_{k=1}^K \gamma(z_{ik}) \left[log\pi_k + log\mathcal{N}\left(x_i \mid \mu_k, \Sigma_k\right)\right] - \lambda(\sum_{k=1}^K \pi_k - 1) \tag{6.11}$$

where λ is a Langrange multiplier. Then, the updated model parameters $\theta^t = \{\mu_k^t, \Sigma_k^t, \pi_k\}$ can be found by evaluating the partial derivative of Eq. 6.11 with respect to each parameter, which yields

$$\pi_k = \frac{\sum_{i=1}^{N} \gamma\left(z_{ik}\right)}{N} \qquad \mu_k^t = \frac{\sum_{i=1}^{N} \gamma\left(z_{ik}\right) x_i}{\sum_{i=1}^{N} \gamma\left(z_{ik}\right)}$$
$$\Sigma_k^t = \frac{\sum_{i=1}^{N} \gamma\left(z_{nk}\right)\left(x_i - \mu_k\right)\left(x_i - \mu_k\right)^T}{\sum_{i=1}^{N} \gamma\left(z_{ik}\right)} \tag{6.12}$$

Once the new parameters are obtained, the process is repeated from the beginning of the expectation step until a convergence criteria is reached.

Initialization

Prior to the first iteration of EM, an initial estimate needs to be obtained for θ. As discussed in Section 2.7.5, θ_{init} can can be estimated be estimated using prior knowledge (even if limited), or initialized using non-informative priors, which is the approach used in this thesis.

6.2.3 Model Selection

The parameter optimization approach described in Section 6.2.2 pertains to a fixed value of K, and a fixed structure for the covariance matrices Σ_k. For the intended applications of this work (i.e. unsupervised CBM), the optimal value of K and the optimal structure of Σ are assumed to be unknown. In this thesis, four main types of structures for the GMM covariance matrices are defined: full-shared, full-unshared, diagonal-shared and diagonal-unshared. Full/diagonal simply refers to whether or not the off-diagonal elements are non-zero (non-zero in the former), while unshared/shared refers to whether or not the covariance matrix should be the same or independent across each Gaussian. In general, the results obtained using the full-unshared covariance should produce the optimal clustering results, as the diagonal and shared options are constraints that are typically applied to improve computational efficiency.

Given that K and the shape of Σ are unknown, selection of the optimal GMM to describe the underlying data requires optimization to be performed with respect to both K and $shape(\Sigma)$. This thesis presents two such approaches to determine these parameters: an information criterion approach using the Bayesian information criterion (BIC); and, the variational Bayesian GMM ($\nu - GMM$) approach.

Akaike and Bayesian Information Criterion

The Akaike and Bayesian information criterion (AIC and BIC, respectively) are model selection criteria based used to evaluate the goodness-of-fit within a set of similar models, as a function of their log-likelihood and model complexity (i.e. number of parameters) [191]. Both AIC and BIC contain a penalty term with respect to the model complexity to discourage overfitting, however, the penalty term in BIC is larger. The smaller penalty term in AIC results in a greater bias towards more complex models relative to BIC, which can lead to overfitting if the underlying (i.e. true) distribution of the data is not complex. Hence, for greater generalization performance in the unsupervised setting, BIC is selected over AIC as the model selection criterion in this thesis. BIC is given by the following expression

$$BIC = klog(N) - 2log(\hat{\mathcal{L}}) \tag{6.13}$$

where k is the total number of model parameters, N is the size of the data, and $\hat{\mathcal{L}}$ represents the maximum log-likelihood. The term $klog(N)$ is the penalty term to control overfitting, and is positive increasing as k increases. Hence, lower BIC values are indicative of a better model fit. Since BIC is not an absolute metric, the important point to make is that BIC (and AIC) are only useful for obtaining a relative comparison of models within the same class (i.e. GMMs with different K), which is the current use case. Given a search range $\mathbb{K} = \{K_1, K_2, ...K_M\}$ for K and four aforementioned types of covariance matrices $\mathbb{C} = \{cov_1, .., cov_4\}$, the optimal GMM model is the one which produces the lowest BIC value across $\mathbb{K}$ and $\mathbb{C}$.

6.2.4 Variational Bayesian Gaussian Mixture Models

Variational Bayesian GMMs (v-GMM) is an extension of the GMM approach which uses variational inference (VI) to simplify the parameter estimation process [132]. In v-GMM, the intractable GMM posterior $p(Z|X, \theta)$ is approximated using a family of distributions $q(Z)$ with simpler functional forms than $p(Z|X, \theta)$. The goal then, is to find the member of $q(Z)$ which gives the best approximation of $p(Z|X, \theta)$. In addition to determining the optimal model parameters θ_{opt}, v-GMM has the additional benefit of implicitly optimizing over $\mathbb{K}$ in the parameter estimation process, which enables K_{opt} to be determined simultaneously. Specifically, v-GMM works by approximating the GMM $p(Z|X, \theta)$ with a mean

field approximation given by

$$p(Z|X) \approx q(Z)$$
$$q(Z) = \prod_{i=1}^{M} q_i(Z_i|X) \tag{6.14}$$

The key simplifying assumption in Eq. 6.14 is that the unknown parameters (i.e. μ,Σ z) are independent. Then, the member of $q(Z)$ that best approximates $p(Z|X)$, can be quantified by

$$q_{opt}(Z) = \arg\min KL[q(Z)||p(Z|X)] \tag{6.15}$$

where KL is the Kullback-Leibler divergence given by

$$KL[q(Z)||p(Z|X)] = \int q(Z) \log \frac{q(Z)}{p(Z|X)} = E\left[\log \frac{q(Z)}{p(Z|X)}\right] \tag{6.16}$$

The KL divergence quantifies the similarity between two probability distributions by measuring the distance between the distributions. However, obtaining q_{opt} from Eq. 6.15 directly is not possible, and hence, the alternative approach is to maximize an equivalent function known as the evidence lower bound (ELBO). It can shown that expansion of Eq.6.16 will yield the ELBO given by

$$\begin{aligned} ELBO(q) = \int q(Z) \log \frac{p(Z,X)}{q(Z)} dZ &= \int q(Z) \log p(Z,X) dZ - \int q(Z) \log q(Z) dZ \\ &= \mathbb{E}[\log p(Z,X)] - \mathbb{E}[\log q(Z)] \end{aligned} \tag{6.17}$$

and q_{opt} is the member of $q(z)$ which maximizes Eq. 6.17.

6.2.5 Coordinate Ascent Variational Inference

In v-GMM, q_{opt} can be solved for using coordinate ascent VI, which is analogous to the previously introduced EM algorithm. Let q^* denote the coordinate update of q. Then, applying the mean field approximation in Eq. 6.14 to $p(z|x)$ and substituting the result into Eq. 6.17 yields the following expression

$$q_j^*(z_j) \propto exp\{\mathbb{E}_{-j}[\log p(z_j|z_{-j}, x)]\} \tag{6.18}$$

where $-j$ denotes all indices other than the j-th. $q_j(z_j)$ is initialized, and then similar to EM, coordinate ascent VI uses the following two step iterative approach to solve for q_{opt}

1. For all $j \in \{1, 2, ..., m\}$, compute Eq. 6.18.
2. Then, compute $ELBO(q)$ using Eq. 6.17.
3. The above steps are repeated until a convergence criteria is reached

6.2.6 Feature Selection

The choice of feature set used for GMM-OSD will affect the decomposition result. In an ideal setting, the selected features for GMM-OSD are those which provide the maximum degree of separation between discrete operating states, which in the supervised setting, can be determined using a variety of feature selection approaches. In the unsupervised setting however, apart from applying bias (i.e. domain knowledge) to pre-select a subset of features that are more likely to be good separators, the optimal set of features cannot be known a priori. In addition, the use of unsupervised ML-based black box feature learning approaches is also challenging, since the number of classes in the underlying data is not known a-priori - and, the features generated by these approaches lack the interpretability of traditional, hand-crafted features.

The proposed feature set consists of eight time and frequency domain features for vibration analysis: RMS, variance, skewness, kurtosis, spectral kurtosis, spectral entropy, mean frequency and half-power bandwidth. The sensitivity of these features to different types of changes in the vibration behaviour of a machine have been well studied in the context of fault detection. The choice of handcrafted features increases the interpretability of the OSD result, the choice of a larger feature set (as opposed to only a few features) provides robustness in the unsupervised use case, which will be demonstrated hereafter in Section 6.3. In addition, a wider feature set provides the GMMs with more measures to differentiate between similar (yet discrete) operating states. Prior to parameter estimation, dimensionality reduction is applied to the feature set using PCA, retaining 95% of the variance.

6.2.7 Validation Approach

Validation of the proposed approach is performed in three steps:

1. First, an exploratory study is conducted to characterize the proposed feature set in the context of OSD across different types of machinery. The purpose here is

qualitatively assess the sensitivity of each individual feature in the feature set to the various operating states across different types of machinery. The proposed feature set is assessed using labelled data from the Pearson LINK APM gearbox and the Pearson Terminal 1 passenger boarding tunnel PCA units. For the LINK APM data, labelled data is used to provide a physical interpretation of the various clusters in the feature data.

2. Next, the classical GMM-OSD approach is applied to the LINK APM and PCA datasets to perform unsupervised OSD. To improve the GMM result on the highly non-stationary LINK APM data, pre-processing using S-MSC is applied to the data prior to GMM-OSD to improve the decomposition result.

3. Finally, the v-GMM approach is applied to both aforementioned datasets to verify that the optimal decomposition results obtained using v-GMM are in agreement with those obtained using GMM and BIC.

6.3 Operating State Characterization

The Pearson Terminal 1 passenger boarding tunnel PCA units and LINK APM gearbox are both examples of multi-operating state machinery. In the case of the former, the discrete operating states of the machine are well defined i.e. four distinct settings to service four different types/sizes of aircraft. This is reflected in both the feature space and resulting GMMs, in which each of the four discrete operating states are well separated from one another. In the case of the LINK APM gearbox however, the operational parameters or conditions which define its different operating states are not as immediately obvious. While the vibration data from the LINK APM certainly does imply the existence of multiple discrete states, the physical phenomena to which those clusters pertain to needed to be determined a-posteriori. The results of this study are presented in **AppendixC**. Using the results of this study, the LINK APM data is labelled and divided into two operating states - each corresponding to one direction of the train's travel. Overall, these two pieces of machinery provide a wide spectrum of operational behaviour to assess the robustness of the proposed feature set.

Given that the proposed approach is to applying GMM clustering, characterization of the feature set across both machines is done with respect to the Gaussian distribution parameters (i.e. mean and variance). For this particular qualitative analysis, a "good" feature can be defined as a feature with well-separated means across each operating state, with similar variance for that feature across operating states. In the context of GMM,

these characteristics will help minimize mixing of individual Gaussians, which will help with convergence during parameter estimation, and reduce uncertainty during posterior inference. The means of the normalized features for the the LINK APM gearbox and Terminal 1 PCA units for each operating state are shown in Fig. 6.2 and 6.3, respectively. Similarly, Fig. 6.4 and 6.5 show the corresponding coefficients of variation (COVs) for each machine.

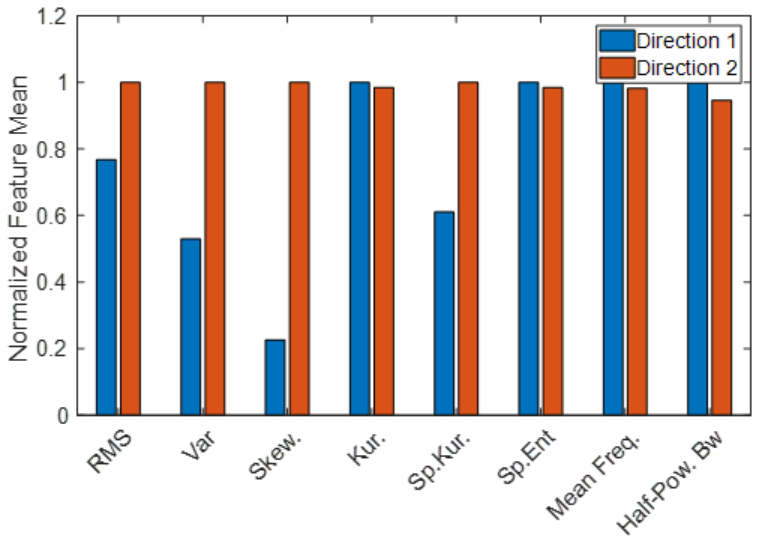

Figure 6.2: Normalized feature means by operating state for Pearson LINK APM gearbox

Evaluation of Fig. 6.2 shows that for the APM, the time domain features (RMS, variance, skewness) and spectral kurtosis contain well-separated means between the two operating states. Considering the COVs of these well-separated features (Fig. 6.4 shows that apart from the skewness, the COVs of the other four features are relatively high, which indicates the presence of non-stationarities within each operating state. In the case of the Terminal 1 PCA units (Fig. 6.3, mean frequency and spectral kurtosis are the best differentiators with respect to the mean. Furthermore, unlike the LINK APM features, the COVs across the PCA feature set are nominal, apart from half power bandwidth, which is indicative of the stationarity of each discrete operating state.

Two key observations can be made from these results. First, the features which provide the highest level of differentiation between the operating states of each machine are different for each machine. In the case of the APM - the best features were the RMS, variance, skewness and spectral kurtosis; while in the case of the PCA units, spectral kurtosis and

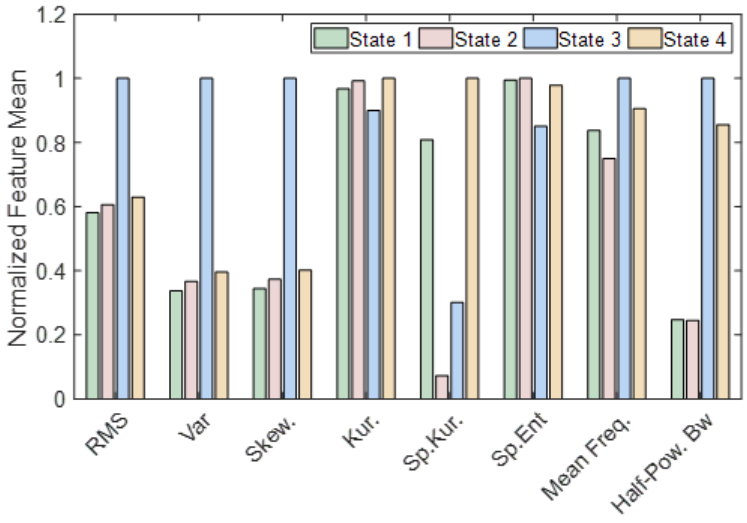

Figure 6.3: Normalized feature means by operating state for Pearson Terminal 1 Passenger Boarding Tunnel PCA Unit

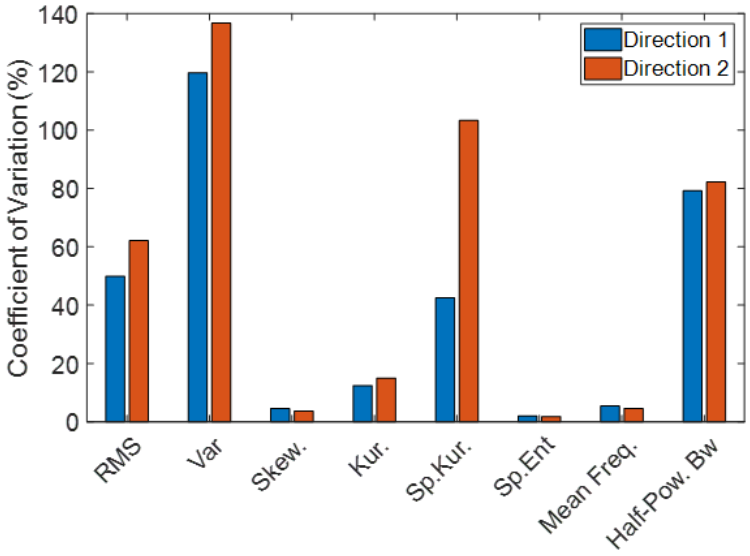

Figure 6.4: Feature coefficients of variation by operating state for Pearson LINK APM gearbox

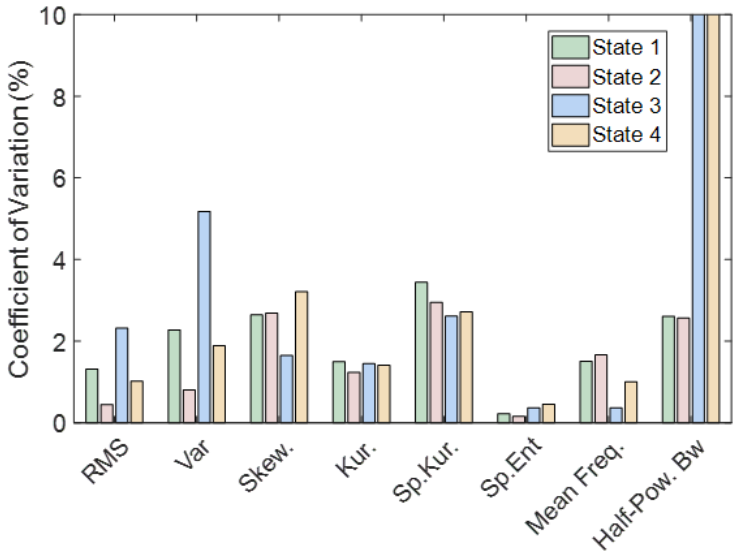

Figure 6.5: Feature coefficients of variation by operating state for Pearson Terminal 1 Passenger Boarding Tunnel PCA Unit

mean frequency were the best separators. For the unsupervised use case, this encourages the use of a broader feature set for robustness. Second, while the APM data contained multiple features with well-separated means, the high COVs resulting from non-stationarity will negatively affect the GMM result. The following section demonstrates how these non-stationarities affect the GMM result, and how pre-processing using S-MSC can be used to non-parametrically remove these non-stationarities and improve the GMM result.

6.4 Application of GMM-OSD with BIC to LINK APM and T1 PCA Unit

This section presents the validation results from the application of GMM-OSD with BIC to the Terminal 1 PCA unit and LINK APM gearbox data. For the LINK APM gearbox data, the GMM results with and without S-MSC pre-processing are presented and discussed.

6.4.1 GMM result on T1 PCA Unit

The GMM-OSD approach is applied to both the Terminal 1 PCA unit and LINK APM gearbox feature data. For each K in $\mathbb{K} = \{1, 2, .., 7\}$ and covariance type, parameter optimization was performed using the EM algorithm, and the best model was selected based on the minimum BIC value. Fig. 6.6 illustrates the BIC and optimal GMM result for the Terminal 1 PCA feature data. Fig. 6.6 shows that for the PCA unit, the four discrete states are easily captured by four, fully separate Gaussians. In the case where each operating state is stationary and clearly differentiable using the proposed feature set, the resulting GMM poses very little uncertainty for posterior interference. In other words, the cluster membership of any unseen point is unambiguous.

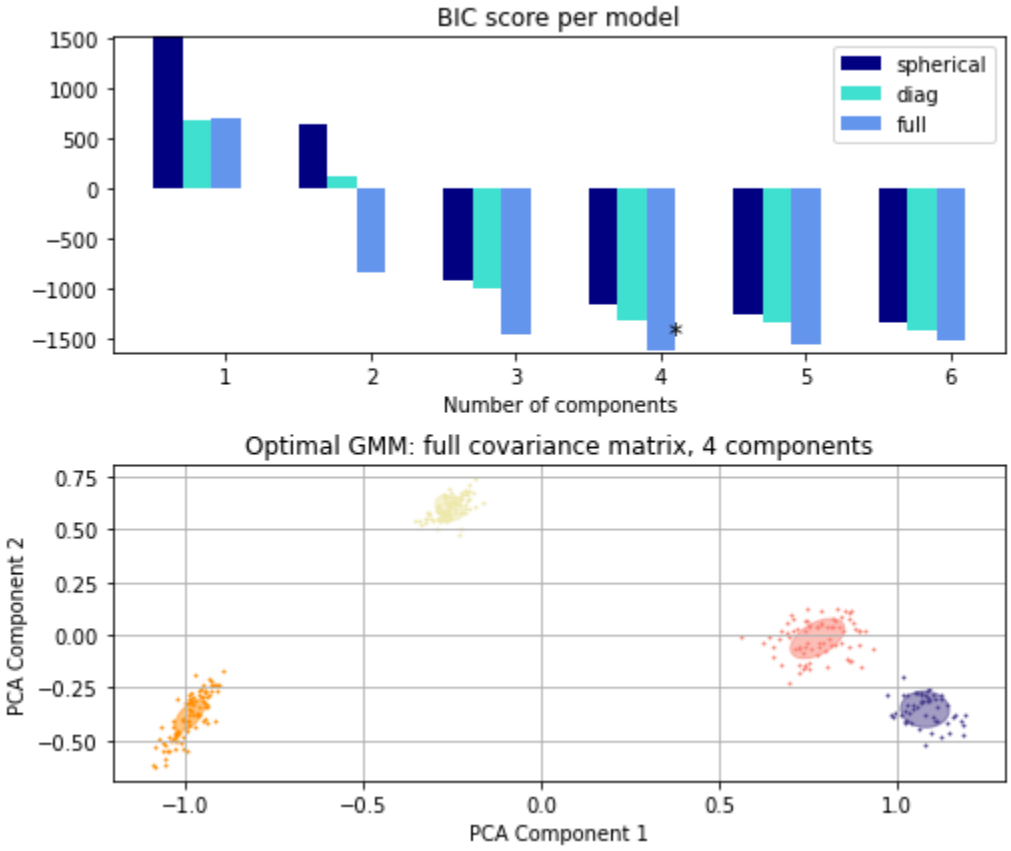

Figure 6.6: Optimal GMM for Terminal 1 PCA Unit (K = 4)

6.4.2 GMM results on LINK APM with S-MSC

The LINK APM gearbox data provides an excellent basis to demonstrate the simultaneous use of two unsupervised pre-processing tools presented in this thesis - S-MSC and GMM-OSD. Fig. 6.7 illustrates the optimal GMM result for the LINK APM data without pre-processing applied. In the case of the LINK APM data, the high variability of the feature data across all operating states produces a convoluted GMM result. In order to reduce the feature variability, S-MSC is applied to the pre-process the raw data and non-parametrically remove the non-stationary components in the signal prior to feature extraction. The process flowchart for this approach was shown earlier in the beginning of the chapter in Fig. 6.1. S-MSC was applied to each windowed segment of the signal to separate the signal into harmonic and residual (noise) signals. The residual signals were discarded, and the resulting harmonic signals were used for feature extraction. The detailed discussion of the application of S-MSC to the LINK APM data can be found in Section 5.5.

Fig. 6.8 illustrates the GMM result with S-MSC applied to pre-process the signal prior to feature extraction. Compared to the un-processed result in Fig. 6.7 for which the lowest BIC value corresponded to a GMM with $K = 6$, the BIC of the S-MSC pre-processed feature data is best described by two Gaussians - which agrees with the posterior knowledge of the operating states present within the LINK APM data.

6.5 Application of v-GMM to LINK APM and T1 PCA Unit

The purpose of this section is to assess the performance of v-GMM-OSD in comparison to the previous GMM-OSD approach with BIC. From the formulation presented in 6.2.4, a primary advantage of v-GMM over traditional GMM is the ability of v-GMM to implicitly optimize over $\mathbb{K}$ during the coordinate ascent VI estimation of the GMM model parameters. To formulate a comparison with traditional GMM, v-GMM is applied in the same manner to both the Terminal 1 PCA unit and S-MSC pre-processed APM gearbox data.

For both cases, v-GMM is initialized with $\mathbb{K} = 10$. Fig. 6.9 illustrates the optimization process over $\mathbb{K}$. At the first iteration, the v-GMM model is initialized using $K = 10$ clusters. As clusters begin to overlap as number of the iterations increases, v-GMM automatically removes overlapping clusters by assigning a weight of zero to redundant clusters. By the 10th iteration, v-GMM has nearly converged to the optimal number of clusters. Fig. 6.10 shows the final, optimal v-GMM result for the S-MSC pre-processed LINK APM gearbox

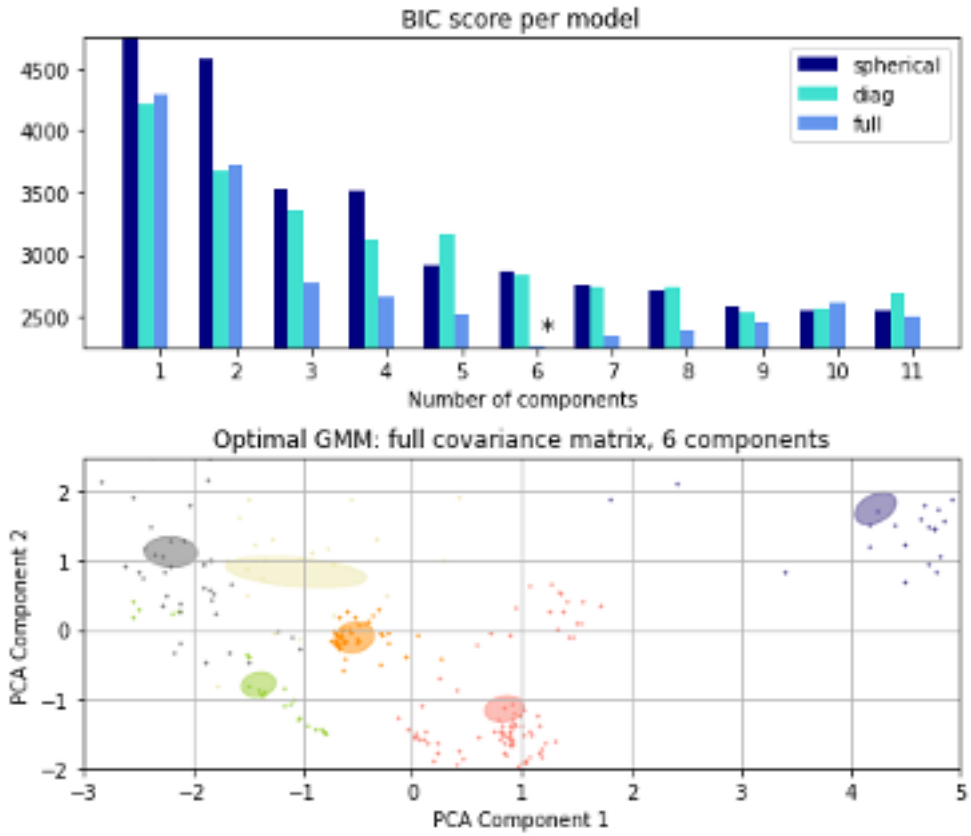

Figure 6.7: GMM result without S-MSC pre-processing - LINK APM Gearbox (K = 6)

data, and comparison of this result to the one obtained using GMM in Fig. 6.8 shows that both results are in agreement. Similar results are obtained when v-GMM is applied to the Terminal 1 PCA data, as shown in Fig. 6.11. The symmetry of these results to those obtained using traditional GMM with BIC demonstrates that v-GMM is a practical improvement over the traditional GMM approach for GMM-OSD, since the model selection with respect to $\mathbb{K}$ is integrated directly with the parameter optimization process. However, the lack of a BIC score in v-GMM means that the result is less interpretable compared to traditional GMM. The BIC in the traditional GMM approach can be used to provide additional insight into the operating behaviour of a machine. For example, the similarity of the BIC values for $K = [2, 3, 4]$ in Fig. 6.8 suggests that the feature data can be clustered to represent multiple sets of physical phenomena beyond train direction.

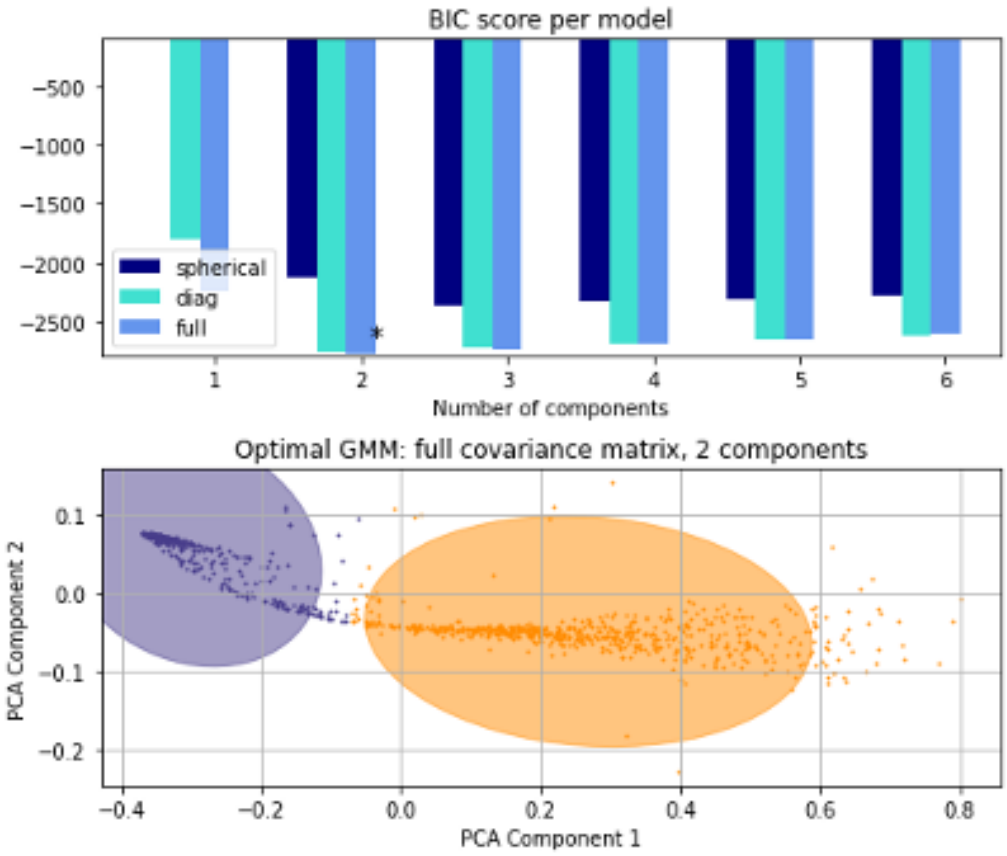

Figure 6.8: Optimal GMM result with S-MSC pre-processing - LINK APM Gearbox (K = 2)

6.6 Summary

This chapter presents an unsupervised pre-processing tool for blind decomposition of vibration signals obtained from multi-modal machinery (GMM-OSD), as well as a combined pre-processing framework in which S-MSC and GMM-OSD are applied in sequence for pre-processing of vibration signals containing multiple types of non-stationarity. For validation, the proposed approach is applied to labelled multi-modal data collected from the LINK APM gearbox and Terminal 1 PCA units for validation. The key conclusions from this work are as follows:

1. A qualitative study was performed on data obtained from the LINK APM gearbox

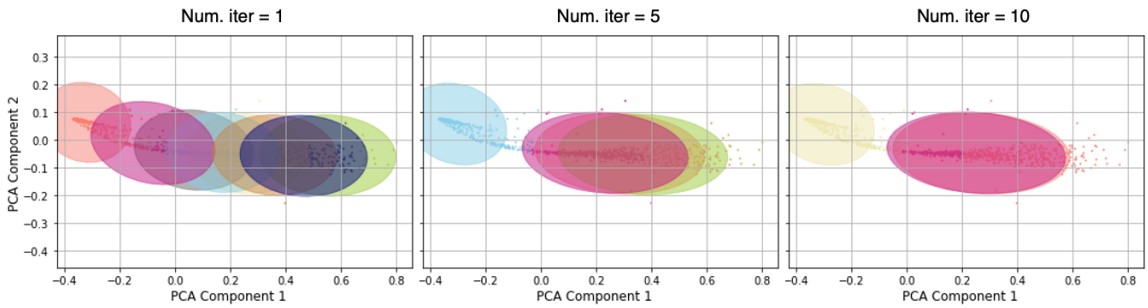

Figure 6.9: Optimization over $\mathbb{K}$ in v-GMM during coordinate ascent VI - LINK APM Gearbox

and Terminal 1 PCA units to characterize the dynamic operational behaviour of these machines and and validate the proposed feature set. The results of the study showed that the best differentiating features varied between the two data sets, which justifies the use of a wider feature set in the unsupervised setting.

2. Applied to data obtained the LINK APM gearbox and Terminal 1 PCA unit field pilots, GMM-OSD was able to blindly decompose the multi-modal data in each signal into the corresponding discrete operating states. For the well-separated Terminal 1 PCA data, GMM-OSD could be applied effectively without pre-processing.

3. For the LINK APM gearbox data, GMM-OSD was unable to cluster the data effectively without pre-processing. After applying S-MSC first as a pre-processing tool to remove non-multi-modal sources of non-stationarity from the signal, GMM-OSD was able to decompose the filtered signal into the correct, discrete operating states.

4. A second formulation using vGMMs was also validated using both datasets, and was shown to converge to the same optimal model selected using BIC in the traditional approach. Both approaches were equally effective for decomposition, with vGMM providing a more computationally efficient solution, and GMM-BIC providing higher interpretability of the model choice due to the use of BIC.

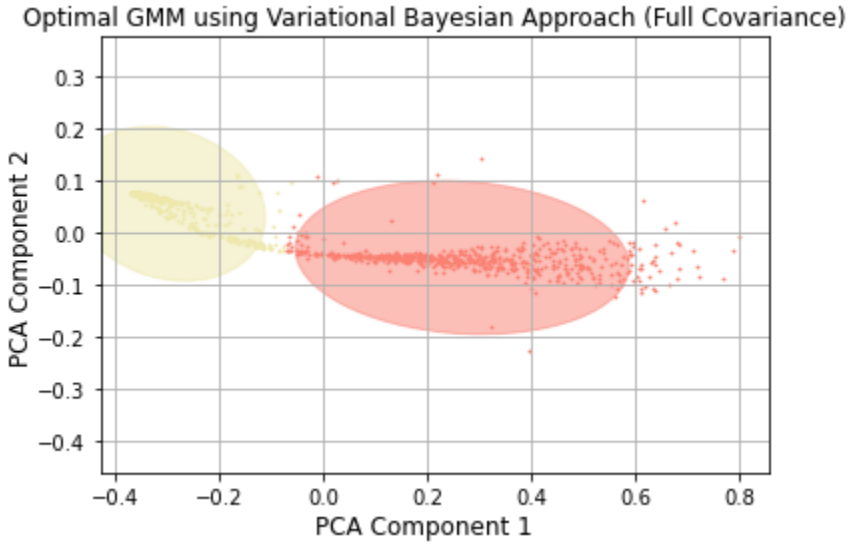

Figure 6.10: Optimal v-GMM of S-MSC Harmonic Features - LINK APM Gearbox (K = 2)

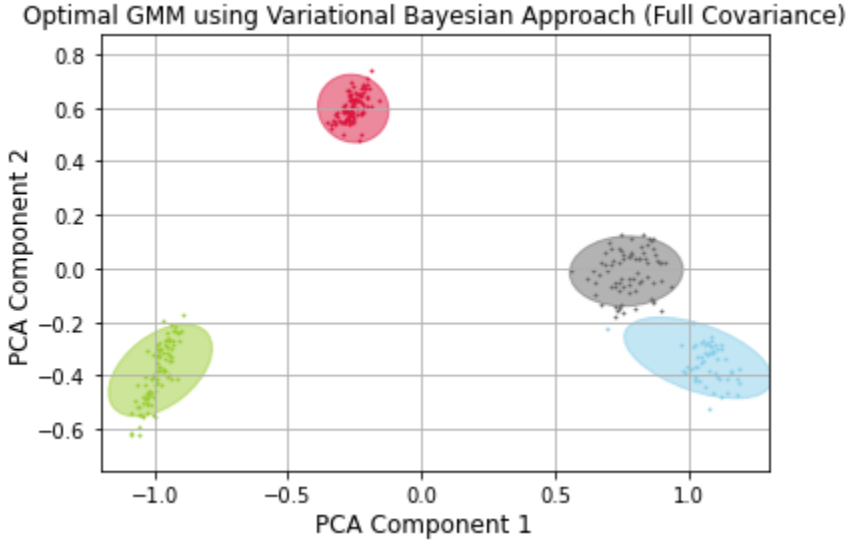

Figure 6.11: Optimal v-GMM for Terminal 1 PCA Unit (K = 4)

Chapter 7

Bayesian One-class Support Vector Machine (B-OCSVM)

This chapter presents an unsupervised approach for early degradation detection and fault prediction, termed the Bayesian One-Class Support Vector Machine (B-OCSVM). The goal of this chapter, and B-OCSVM, is to tackle the several key gaps in literature pertaining to the application of unsupervised CBM in the real world - namely, the issues of hyperparameter estimation for anomaly detection and degradation threshold setting in the absence of prior knowledge. Ultimately, these key knowledge gaps severely hinder the applicability of these CBM approaches in real world, unsupervised settings.

The chapter begins by reviewing the specific gaps in knowledge and motivations behind the approach, followed by an overview and detailed description of the methodology. The final sections of this chapter presents results obtained from three public run-to-failure datasets, which are used for validation, benchmarking, sensitivity analysis, and evaluation of robustness and generalization performance.

7.1 Motivations for B-OCSVM

While a detailed discussion on the current state-of-the-art and overview of the key knowledge gaps within unsupervised CBM methods are presented in Sections 3.3.3 and 3.5, respectively, the following summary is intended to provide the reader with the necessary context required for understanding the motivations behind the work presented hereafter. The key observations from the current-state-of-the art are as follows:

1. The majority of unsupervised and semi-supervised CBM approaches employ some form of support vector-based classifier for fault detection (i.e. OC-SVM, SVDD, or N-SVDD). This supports the choice of type of classifier for the current application.

2. Within the current body of literature, however, the challenging task of unsupervised hyperparameter estimation has not been adequately addressed. Additionally, existing approaches for unsupervised OC-SVM hyperparameter estimation are predicated on assumptions placed on the data which cannot be easily realized in the real-world setting. Hence, there currently is no established framework available to facilitate the implementation of the state-of-the-art to unsupervised applications in the real-world.

3. Similarly, methods for determining a suitable threshold for degradation modeling are largely dependent upon prior knowledge (i.e. failure records, previous fault observations). Even in unsupervised approaches, there is insufficient guidance pertaining to the estimation of a suitable threshold without the use of prior information or posterior inference. Without a suitable degradation threshold, these methods cannot be applied effectively in the unsupervised setting.

4. Likely related to the previous point is the fact that many existing unsupervised approaches are heavily focused on the detection aspect of CBM, and lacking in prognostic capabilities, which are often of equal practical significance. Several studies have proposed the use of ML classifier distance from the decision boundary as a means to infer the severity of degradation, but none of these studies have provided a formal framework applicable to the unsupervised setting.

5. Finally, while not as large of an issue as the aforementioned gap areas, the vast majority of methods rely on features are extracted using black-box ML processes meaning that the interpretability of these features is low relative to traditional handcrafted features. The well-understood behaviour of traditional handcrafted features with respect to different types of machinery faults can provide some intrinsic diagnostic value in an otherwise information-scarce setting.

7.2 Background

B-OCSVM is an unsupervised approach for early degradation detection and fault prediction based on the OC-SVM [166] and the hierarchical Bayesian framework [52]. Unlike traditional CBM frameworks which aim to detect well-developed faults, B-OCSVM instead shifts the focus towards the detection of the degradation process earlier on in the

fault development cycle, and aims to predict the time until a detected degradation process will manifest into a well-developed fault. In other words, the intent of B-OCSVM is fault prediction, as opposed to the traditional approach of fault detection failure prediction. Fig. 7.1 illustrates the idea of early degradation region targeted by B-OCSVM in relation to the typical fault detection region found in traditional fault detection approaches.

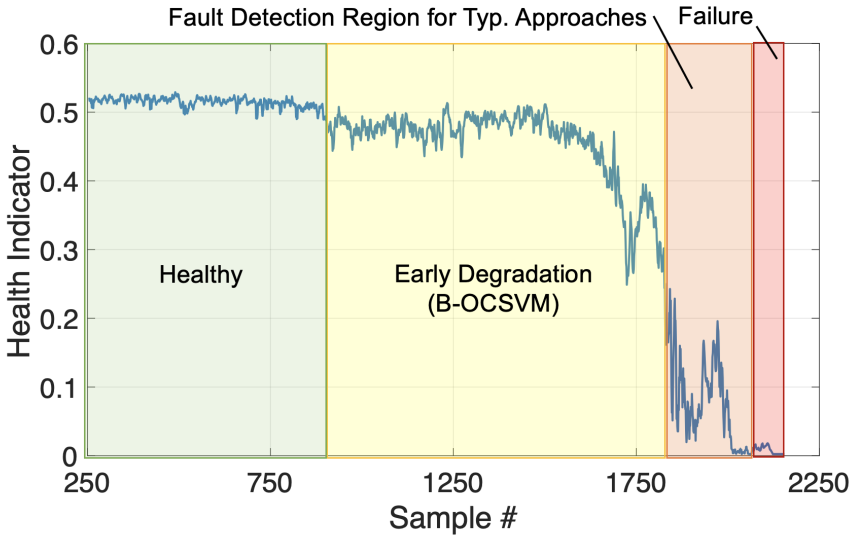

Figure 7.1: Early degradation region targeted by B-OCSVM vs. traditional fault detection region for IMS bearing S1B3 [101]

The overarching goal of B-OCSVM to provide a fault prediction framework that can be easily translated into real-world unsupervised settings. This requires the ability to estimate both the OC-SVM hyperparameters and degradation thresholds without the reliance on prior knowledge. Hence, new approaches to determine both of these parameters in the unsupervised setting are proposed. In B-OCSVM, early detection of the degradation process and fault prediction are achieved using the scaled logistic sigmoid output of the OC-SVM classifier in tandem with a prior-independent degradation threshold that is defined as a function of the classifier decision space. The task of estimating the optimal values of OC-SVM hyperparameters γ and ν in the unsupervised setting is achieved using a novel parameter estimation approach which hybridizes heuristics and brute-force parameter optimization. When a degradation process is detected by B-OCSVM,a hierarchical Bayesian degradation framework is used in conjunction with a novel prognostic, termed time-to-threshold (t^3), to predict the time at which the degradation process will manifest

into a fault.

7.3 Methodology

This section presents the specific background and detailed methodology for the B-OCSVM approach. A summary of the key steps in the methodology is provided at the end of the section.

7.3.1 Basic Assumptions

In order to satisfy the conditions for unsupervised learning, only unlabelled data is used for training with a single imposed bias [115] that the training data is sampled from the healthy operating state of a machine and can be contaminated with some unknown fraction of anomalies ρ.

7.3.2 Feature Extraction

The handcrafted features used in B-OCSVM consists of well-studied set of traditional diagnostic features from the time, frequency and time-frequency domain: root-mean square, variance, skewness, kurtosis, shape factor, absolute energy, Shannon entropy, spectral entropy, spectral kurtosis and wavelet entropy [186]. To further demonstrate the robustness of the feature set in the unsupervised setting, no traditional pre-processing steps, such as de-noising or demodulation, are performed on the signals prior to feature extraction. All features are extracted providing only the sampling frequency f_s as an input argument. Prior to training the OC-SVM classifier , dimensionality reduction is applied to the feature set using PCA [196] (retaining 95% of the variance) in order to prevent the curse of dimensionality.

7.3.3 OC-SVM

The OC-SVM classifier is a special case of SVM. Recall from section that the primary goal in SVM is to transform a dataset into a higher dimensional space that results in the maximal margin decision boundary between the different classes in the dataset. As the name implies, OC-SVM is simply the special case of SVM in which there is only one class contained in the

data. Hence, the goal in OC-SVM is to provide the maximal margin of separation between the data and the origin. Specifically, given a training set $X = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^d$, the OC-SVM classifier finds a decision boundary that encapsulates the densest regions of X [166]. A test point x is classified as normal if it lies within the boundary, and anomalous otherwise. Often times, a transformation of the data to a higher dimensional space (i.e. $\varphi : \mathbb{R}^d \to \mathbb{R}^{d'}$, where $d' \ggg d \in \mathbb{R}^{d'}$) is applied to make the data linearly separable, resulting in the primal quadratic problem [166]

$$\begin{aligned} \min_{\forall \omega, \xi, \rho} \frac{1}{2}\|\omega\|^2 - \rho + \frac{1}{\nu N}\sum_{i=1}^{n}\xi_i \\ \text{s.t. } \nu \in (0,1],\ \langle \omega \cdot \varphi(x_i) \rangle \geq \rho - \xi_i \\ \xi_i \geq 0, \forall i = 1, 2, ..., n \end{aligned} \tag{7.1}$$

where $\omega \in \mathbb{R}^d$, ξ_i are slack variables, ρ is the variable which represents the bias, and ν is a hyperparameter which dictates the upper bound on the fraction of training errors (i.e. points that will be classified as an anomaly) and the lower bound on the number of training points that can act as SVs. Evaluating the limit as ν approaches zero results in the slack variables $\xi_i \to 0$, and a hyperplane that separates nearly all of the training data from the origin in $\mathbb{R}_{d'}$. However, forcing ν towards zero proportionally increases risk of overfitting on the training data. In Eq. 7.1, the distance from the origin to the hyperplane is given by $d_{hyper} = \frac{\rho}{\|\omega\|}$, and the maximal solution for d_{hyper} is obtained by minimizing the term $\frac{1}{2}\|\omega\|^2$. For computational efficiency, the kernel trick with kernel function $K(x, x_i) = \varphi(x)^T\varphi(x_i)$ is used to compute φ implicitly, from which Eq. 7.1 can be re-written as the following quadratic dual problem:

$$\begin{aligned} \min_{\forall \alpha_i} \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j K(x_i, x_j) \\ \text{where } K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2} \\ \text{s.t. } 0 \leq \alpha_i \leq 1/\nu n \\ \sum_{i=1}^{n}\alpha_i = 1, \forall i = 1, 2, ..., n \end{aligned} \tag{7.2}$$

where α_i are the Langrange multipliers. In this form, $\rho = \sum_j \alpha_j K(x_j, x_{sv})$, where x_{sv} represents any arbitrary SV. In this approach, $K(x, x_i)$ is the widely adopted Gaussian kernel described in section 2.30. The corresponding decision function of 7.2 is then given

by:

$$f(x) = sgn(\sum_{j=1}^{nsv} \alpha_j K(x, x_j) - \rho)$$
$$\text{where } f(x) = \begin{cases} \text{anomaly if } f(x) = -1 \\ \text{healthy if } f(x) = 1 \end{cases} \tag{7.3}$$

From Eq. 7.2 and 7.3, the signed distance of a new data point x from the separating hyperplane can be obtained as

$$dist(x) = \frac{\sum_{j=1}^{nsv} \alpha_j K(x, x_j) - \rho}{\|\omega\|} \tag{7.4}$$

7.3.4 OC-SVM Hyperparameter Estimation for Early Degradation Detection

Early detection of the degradation process requires choosing values of γ and ν that minimize the volume of the training set (i.e. tight decision boundary). Recall from section 7.3.3 that while decreasing the value of ν will result in a tighter decision boundary, there is an inherent risk or tradeoff in regards to overfitting. An additional practical constraint is that the parameter estimation should converge even on small training sets, as the large volumes of data required in many existing hyperparameter approaches may not be easily obtainable in practice.

Existing unsupervised OC-SVM hyperparameter estimation methods can be classified as either heuristic [54, 114] or grid-search based [154, 175] approaches. Existing heuristic approaches, while efficient, may under-perform or overfit when the data in question does not satisfy underlying assumptions or when the training set is small. Grid-search approaches can circumvent the aforementioned pitfalls of heuristic methods through careful objective function design, but these approaches suffer primarily from high computational cost, particularly when both hyperparameters are solved for using this manner.

The motivations behind the proposed OC-SVM hyperparameter approach are best understood using an example. Consider the KDES of the 2-D PCA projection of a healthy bearing (IMS bearing S1B3 [101]) and a bivariate standard normal random variable ($Z(x, y)$) shown in Fig. 7.2 illustrates the KDEs. Both sets of data contain the same number of points (n = 425), and can be described by the same first and second order statistical moments. However, the KDE of bearing S1B3 is far more complex than that

of $Z(x, y)$, consisting of several lower density regions interspersed between higher density regions. The corresponding OC-SVM decision boundary for the IMS bearing data fitted using the proposed hyperparameter estimation approach and a density-based heuristic approach are shown in Fig. 7.3. Using the heuristic approach shown on the right in Fig. 7.3, the complex density distribution of S1B3 results in the classifier overfitting on the data, while the decision boundary determined using the proposed estimation approach (Fig. 7.3 left) is able to encapsulate the training data with the minimal margin of separation without overfitting in the process. Therefore, the decision boundary obtained using the proposed hyperparameter estimation approach enables an underlying degradation process to be detected early on while preventing a high false positive rate.

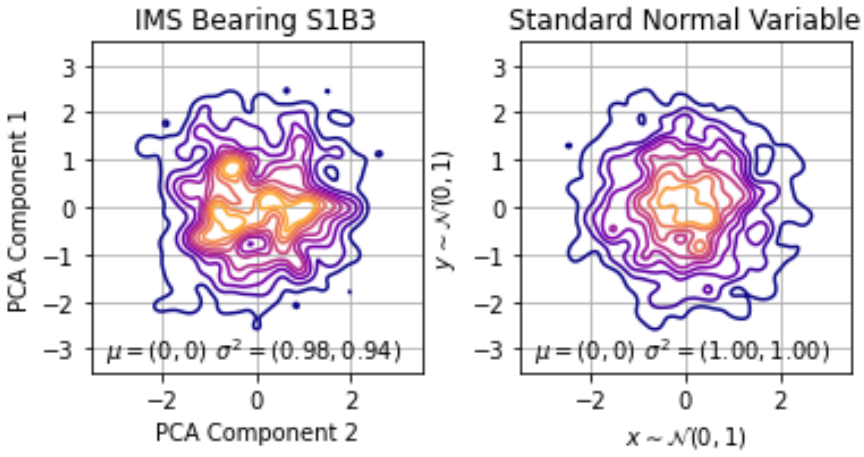

Figure 7.2: Kernel density estimate for: bearing vibration PCA data (left) and bivariate standard normal variable (right)

7.3.5 Proposed OC-SVM Hyperparameter Estimation Procedure

The proposed approach aims to circumvent the aforementioned pitfalls of existing OC-SVM hyperparameter estimation method by using a hybridized heuristic and optimization-based approach. First, a heuristic approach to filter the training data and set ν, followed by the estimation of γ using a newly defined objective function in conjunction with grid search optimization. This hybrid approach leverages computational efficiency of heuristic approaches while preventing the overfitting problem in these approaches shown in Fig. 7.3 through the use of grid-search optimization. Furthermore, the proposed approach satisfies

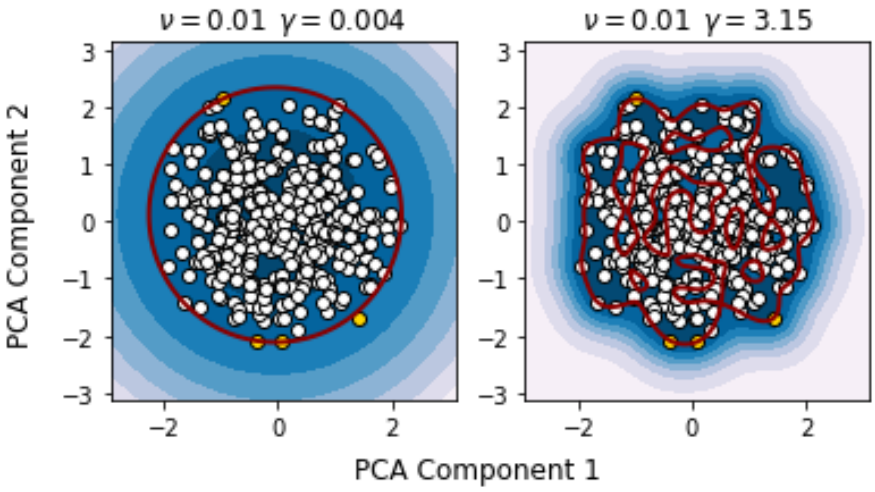

Figure 7.3: OC-SVM decision boundary for IMS bearing S1B3 using: proposed parameter estimation approach (left) density-based parameter estimation approach (right)

the additional practical constraint by allowing the OC-SVM classifier to find the optimal decision boundary even when the size of the training set is small. The integration of a heuristic filter the training data and to set ν drastically reduces the complexity of the estimation approach from $O(N^2)$ to $O(N)$.

The proposed OC-SVM hyperparameter estimation approach is comprised of three main steps: (1) outlier removal, (2) heuristic estimation of ν and (3) grid search optimization-based estimation of γ.

Outlier Removal

OC-SVM is highly sensitive to the presence of outliers in the training data as these points can easily be misidentified as support vectors by the classifier in the unsupervised setting [62, 4], which can result in delayed detection of the degradation process in the context of CBM. In order to reduce their influence on the decision boundary and increase the sensitivity of the classifier to early degradation, outliers must be removed from the training set prior to parameter estimation [62].

Outlier removal is performed using a k-Nearest Neighbours (k-NN) based filtering approach for its strong performance in detecting global outliers without over-sensitivity towards local outliers [107]. The $L2$ k-NN distance $d_k^{L2}(x_i)$ of point x_i from its k nearest

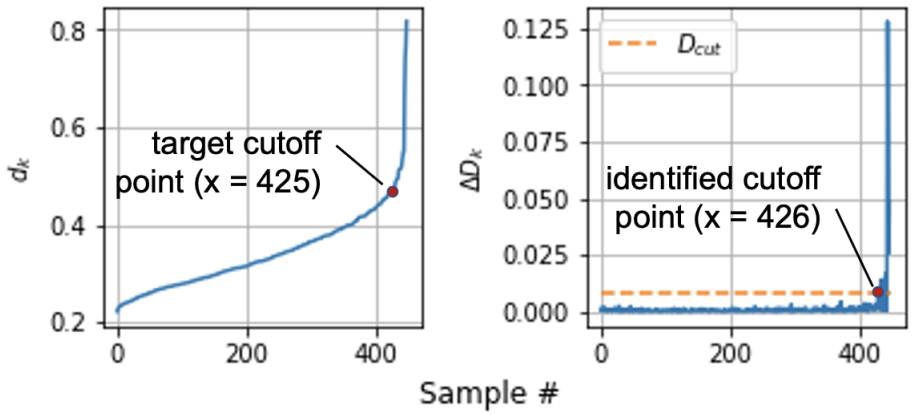

Figure 7.4: k-NN distances for IMS bearing S1B3: sorted ascending distances D_k (left) and 1^{st} difference of D_k (ΔD_k) with proposed outlier threshold D_{cut} (right)

neighbours is given by:

$$d_k^{L2}(x_i) = \frac{1}{k}\sqrt{\sum_{j=1}^{k}(x_i - x_j)^2} \tag{7.5}$$

In Eq. 7.5, the k-NN error asymptotically approaches a minimum as $k, n \to \infty$ and $k/N \to 0$. Hence, $k = 5log(n)$ is set to satisfy this condition [107]. In the training set, d_k is minimized for points residing belonging to a high density region, and maximized for points farthest from these regions. Hence, points with the largest values of d_k can be considered as potential outliers [56]. A key challenge in the unsupervised setting is that the exact fraction of outliers in the training set, denoted by η, is not known a priori. Existing approaches for determining a suitable outlier threshold requires either prior knowledge of η or the use of additional expensive algorithms [54, 114]. In response, a new computationally efficient method is proposed to determine a suitable threshold for k-NN-based outlier removal.

Let D_k represent the sorted ascending list of $d_k(x_i)\ \forall\ x_i \in X$. D_k for IMS bearing S1B3 is shown on the left in Fig 7.4. The constant slope region of D_k for S1B3 (i.e. the range between sample 0 to 425) corresponds to the data within the high density regions, while the remaining region to the right of this data corresponds to outliers. Hence, the changepoint between these two regions (x = 425) can be used as a cutoff point for outlier removal [54]. To determine this changepoint in an unsupervised setting, the following efficient method

is proposed, using the first difference ΔD_k of D_k, given by:

$$\begin{aligned} \Delta D_k &= \{\Delta D_k(1), \Delta D_k(2), ..., \Delta D_k(n)\} \\ \text{where } \Delta D_k(i) &= D_k(i) - D_k(i-1) \\ &\forall i = 1, 2, .., n \end{aligned} \tag{7.6}$$

ΔD_k for IMS bearing S1B3 is shown on the right in Fig 7.4. According to the behaviour of D_k, ΔD_k is stationary for all non-outlier points, and increasing for outliers. A closed form solution to determine the changepoint (i.e threshold) D_{cut} can be derived using the weighted mean μ_w and weighted standard deviation σ_w of ΔD_k as follows:

$$\begin{aligned} D_{cut} &= \mu_w + \sigma_w \\ \text{where } w_i = 1 - \frac{|\mu - \Delta D_k(i)|}{\max|\mu - \Delta D_k|}, \quad \mu_w &= \frac{\sum_{i=1}^{n} w_i \Delta D_k(i)}{\sum_{i=1}^{n} w_i} \\ \sigma_w &= \left[\frac{\sum_{i=1}^{n} w_i (\mu_w - \Delta D_k(i))^2}{\sum_{i=1}^{n} w_i}\right]^{0.5} \end{aligned} \tag{7.7}$$

In Eq. 7.7, μ is the unweighted mean of ΔD_k. A point x_i with $\Delta D_k(i)$ exceeding D_{cut} is classified as an outlier and removed from the training set. The cutoff point identified by proposed method (x = 426; Fig. 7.4a) is in close agreement with the targeted cutoff point (x = 425; Fig. 7.4b). The role of the weights w_i is to reduce the influence of outliers when calculating the sample statistics, which is particularly important in cases where n is small or when η is non-trivial.

Heuristic estimation of ν

Prior to outlier removal, the fraction of outliers η in the training set is unknown in the unsupervised setting. However, removal of outliers from the training set places an implicit assumption on η (i.e. $\eta \approx 0$ after filtering), which enables ν to be efficiently determined using a heuristic [54]. It was shown in [54] that following outlier removal, the resulting decision boundary is robust to the choice of ν. Hence, for consistency with [54], $\nu = 0.01$.

Estimation of γ using grid-search optimization

To estimate γ, the following new objective function is defined based on the heuristic used to determine ν, which is evaluated using 5-fold cross validation:

$$\arg\min_{\gamma} J(\gamma) = \frac{[n - \sum_{i=1|f(x_i,\gamma)=-1}^{n} f(x_i,\gamma)] - (1-\nu)n}{(1-\nu)n} \quad \text{where} \;\; \gamma = 2^m, m \in \mathbb{Z} : m \in [-18, 2] \tag{7.8}$$

where n is the number of the points in the filtered training set, and $\sum_{i=1|f(x_i,\gamma)=-1}^{n} f(x_i,\gamma)$ is the sum of all training samples classified as an anomaly for a given value of γ, $(1-\nu)n$ is the theoretical fraction of normal points in the filtered training set. In Eq. 7.8, the optimal value of γ is the one that minimizes the difference $[n - \sum_{i=1|f(x_i,\gamma)=-1}^{n_t} f(x_i,\gamma)] - (1-\nu)n$, which thereby satisfies the heuristic used to determine ν. To prevent overfitting, Eq. 7.8 is evaluated using 5-fold cross validation over a typical OC-SVM γ search range.

7.3.6 Logistic Sigmoid Normalized OC-SVM Distance as a Degradation Surrogate

For most fault detection applications, the binary result obtained using decision function in Eq. 7.3 is often sufficient. However, as proposed by several other unsupervised and semi-supervised works, the raw classifier output (i.e. signed distance) can be used to infer the severity of the underlying degradation process [129, 169, 39]. This is true for the case of SVM and OC-SVM - the distance of a data point from the decision boundary can be interpreted as an implicit measure of severity. Hence, the proposed approach uses a scaled version of the OC-SVM output is used as a surrogate to infer the severity of the underlying degradation process. Prior to detection and prediction, the signed OC-SVM distance is normalized through convolution with a logistic sigmoid function. The logistic sigmoid $S(x)$ maps an input onto an S-curve bounded between 0 and 1, and is given by:

$$S(x) = \left[1 + e^{-\zeta x}\right]^{-1} \quad \text{where} \;\; 0 < S(x) < 1, \; \forall x \in \mathbb{R} \tag{7.9}$$

where ζ is a scale parameter that controls the steepness of the curve. Due to the exponential e^{-x}, saturation of Eq. 7.9 can be achieved over a small range of x (i.e. $[x^{-}_{lim}, x^{+}_{lim}] = [-6, 6] \in \mathbb{R}$. The logistic sigmoid with $\zeta = 1$ over $x \in [-6, 6]$ is shown in Fig. 7.5.

$S(x)$ provides several benefits when applied as the convolution kernel to the OC-SVM output. $S(x)$ magnifies monotonic degradation trends within the subset of points close to the decision boundary i.e. $d(x) \approx 0$. Furthermore, convolution with the sigmoid function

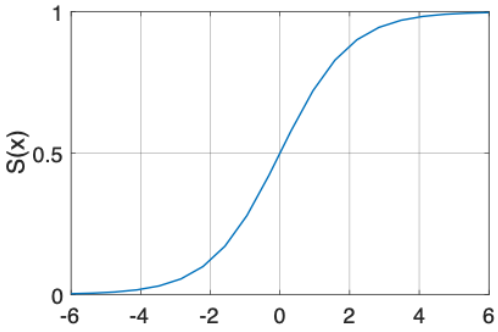

Figure 7.5: Logistic sigmoid function

standardizes the input to the range $(0, 1)$, which can beneficial when combining multiple classifiers in ensemble learning applications. Finally, and most importantly, the horizontal asymptotes of $S(x)$ provide natural boundaries that can be scaled accordingly to define suitable thresholds for degradation modeling and fault prediction. In this approach, a scaling factor is proposed for the steepness parameter ζ in Eq. 7.9 that integrates the margin $\frac{\rho}{\|\omega\|}$ of the OC-SVM classifier into $S(x)$, from which a new prior-independent degradation threshold can be defined. The resulting scaled logistic sigmoid transformation, denoted by $S_{svm}(x_{new})$ is given by:

$$S_{svm}(x) = \left[1 + exp\Big(\frac{-\zeta}{\|\omega\|}\sum_{j=1}^{nsv}\alpha_j K(x, x_j) - \rho\Big)\right]^{-1} \qquad \text{where } \zeta = x^{+}_{lim}/\rho \tag{7.10}$$

where the parameter $\zeta = x^{+}_{lim}/\rho$ ensures that the scale of the original OC-SVM decision space is preserved in the new transformed space. The OC-SVM decision function in Eq. 7.3 can now be expressed in terms of the transformed variable S_{svm}:

$$f(S_{svm}(x)) = sgn(0.5 - S_{svm}(x)) \tag{7.11}$$

where a $f(S_{svm}(x)) = -1$ indicates an anomaly.

7.3.7 OC-SVM-Scaled Sigmoid Degradation Threshold

Using Eq. 7.10, a new prior-independent degradation threshold for fault prediction is derived. Recall from the previous section that in the OC-SVM decision space $\mathbb{R}^{d'}$, anomalous

points are situated in the region bounded by the decision boundary and the origin (i.e. $0 < \varphi(x) < \frac{\rho}{\|\omega\|}$). Points further from the decision boundary imply a higher level of severity, with points situated at the origin representing the maximum deviation from the healthy class (i.e. indicative of a fault). The scaling applied to ζ in Eq. 7.10 means that the lower sigmoid asymptote of $S_{svm}(x)$ is analogous to the origin in the $\mathbb{R}^{d'}$ OC-SVM decision space. Hence, the lower asymptote of Eq. 7.10 is proposed as a new degradation threshold for fault prediction. To account for the asymptotic behaviour of S_{svm}, a nominal non-zero value is chosen and the degradation threshold $T_{S_{svm}}$ is defined as the following:

$$T_{S_{svm}} \equiv \lim_{x \to -\infty} S_{svm} \approx S_{svm} = 0.05 \tag{7.12}$$

The results shown in section 7.6 will validate that the degradation threshold proposed in Eq. 7.12 is indicative of the level of degradation at which other approaches typically detect a fault.

7.3.8 Smoothing and Degradation Detection Criteria

To minimize the false positive rate while monitoring, two post-processing measures are applied to the normalized OC-SVM output S_{svm} from Eq. 7.10 prior to classification. First, S_{svm} is passed through a moving-average filter:

$$\bar{S}_{svm}(w_k) = \frac{1}{m} \sum_{j=k-m+1}^{k} S_{svm_j} \tag{7.13}$$

where w_k is the k-th sliding window of size m. For all the experimental validation performed in this thesis, $m = 5$ is used. Finally, a moving window majority vote decision criterion based on the fault detection criteria proposed in [124] is applied to filtered output, resulting in the final decision function:

$$df(\bar{S}_{svm_k}) = \begin{cases} \text{anomaly} & \text{if mode}[f(w_{\bar{S}_{svm_k}})] = -1 \\ \text{healthy} & \text{otherwise} \end{cases}$$
$$\text{where } w_{\bar{S}_{svm_k}} = \left[\bar{S}_{svm_k}, \bar{S}_{svm_{k-1}}, ..., \bar{S}_{svm_{k-l+1}}\right], l = 5 \tag{7.14}$$

7.3.9 Hierarchical Bayesian Degradation Modeling

Fault prediction is performed using a Hierarchical Bayesian degradation modeling approach, which is based off the hierarchical Bayesian framework described in section 2.7.5.

In this application, the Bayesian framework provides a suitable, general solution to the unsupervised problem where model parameters cannot be assumed to be known prior to the start of monitoring [52]. In this approach, the hierarchical Bayesian framework is applied to the exponential degradation model, and initialize each model with non-informative priors. Then, as new observations become available, the true priors are obtained using hierarchical Bayesian updating.

Exponential Degradation Model

The exponential degradation model is selected for four reasons: first, the exponential degradation model has been used extensively to model the degradation of many mechanical components [52]. Second, the scaled logistic sigmoid normalization applied the OC-SVM output (Eq. 7.10) imposes an exponential or linear functional form on the data. Third, for practical considerations, model updating for the exponential model is computationally inexpensive since a closed-form solution exists for the joint-posterior of the model parameters. Finally, it is assumed that since only a single phase (i.e. incipient) of the degradation process is considered, the degradation behaviour can be sufficiently modeled using a single Bayesian process. The exponential degradation model assumes that the state of degradation at time k, denoted by Y_{t_k} is given as:

$$Y(t_k) = \phi + \theta e^{\beta t_k + \epsilon(t_k) - \frac{\sigma^2}{2}} \tag{7.15}$$

where ϕ is a constant, $\ln\theta \sim N(\mu_\theta, \sigma_\theta^2)$ and $\beta \sim N(\mu_\beta, , \sigma_\beta^2)$ are normal random variables and $\epsilon(t_k) \sim N(0, \sigma^2)$ is standard normal error term. For mathematical convenience, Y_{t_k} is often expressed in the logarithmic form $\mathcal{L}_{t_k}$:

$$\begin{aligned} \mathcal{L}(t_k) &= \theta' + \beta(t_k) + \epsilon(t_k) \\ \text{where } \mathcal{L}(t_k) &= \ln(Y_k - \phi) \\ \theta' = ln\theta - \sigma^2/2 &\sim N(\mu_{\theta'}, \sigma_\theta^2) \\ \mu_{\theta'} &= \mu_\theta - \sigma^2/2 \end{aligned} \tag{7.16}$$

The joint posterior distribution of θ and β follows a bivariate normal distribution with mean $(\tilde{\mu}_\theta, \tilde{\mu}_\beta)$ and variance $(\tilde{\sigma}_\theta^2, \tilde{\sigma}_\beta^2)$ from which the mean and variance of the degradation process at time t in the future, estimated from the current time t_j, is given by [52]:

$$\begin{aligned} \tilde{\mu}_{\mathcal{L}_{t+t_j}} &= \tilde{\mu}_{\theta'} + \tilde{\mu}_\beta(t + t_k) - \frac{\sigma^2}{2} \\ \tilde{\sigma}^2_{\mathcal{L}_{t+t_j}} &= \tilde{\sigma}^2_{\theta'} + \tilde{\sigma}^2_\beta(t + t_k)^2 + \sigma^2 \end{aligned} \tag{7.17}$$

7.3.10 A New Prognostic: Time-to-Threshold (t^3)

In conjunction with the previously defined degradation threshold T_{svm}, a new prognostic called the time-to-threshold (t^3) is defined, which represents the probabilistic estimate of the time until the degradation process manifests into a fault at the degradation threshold $T_{S_{svm}}$. From Eq. 7.17 the cumulative density function (CDF) of t^3 at time step k is given by:

$$\begin{aligned} t_k^3 &= P(L_j < t|\mathcal{L}_1, ..., \mathcal{L}_j) \\ &= P(\mathcal{L}_{t+t_k} > T_{S_{svm}}|\mathcal{L}_1, ..., \mathcal{L}_j) \\ &= \Phi\Big(\frac{T_{S_{svm}} - \tilde{\mu}_{\mathcal{L}_{t+t_j}}}{\tilde{\sigma}_{\mathcal{L}_{t+t_j}}}\Big) \end{aligned} \tag{7.18}$$

where $\Phi(\cdot)$ is the CDF of the standard normal random variable Z.

7.3.11 Summary of Methodology

The key steps in the B-OCSVM approach are summarized as follows:

1. Training:
   (a) Feature extraction and dimensionality reduction
   (b) OC-SVM hyperparameter estimation:
      i. Remove outliers using NN-based approach (Eq. 7.5-7.6) with proposed outlier threshold (Eq. 7.7).
      ii. Set $\nu = 0.01$ heuristically.
      iii. Estimate γ through k-fold CV grid search optimization of the proposed objective function $J(\gamma)$ (Eq. 7.8)
   (c) Train OC-SVM classifier using obtained optimal hyperparameters
   (d) Compute proposed fault prediction threshold $T_{S_{svm}}$ (Eq. 7.12)
2. Online monitoring (degradation detection):
   (a) Compute S_{svm} using the scaled-logistic sigmoid convolution (Eq. 7.10)
   (b) Smoothing of S_{svm} (Eq. 7.13)

(c) Detection of the incipient degradation process using decision function $df(\bar{S}_{svm_k})$ (Eq. 7.14)

3. Online monitoring (fault prediction):

   (a) Upon detection of degradation process, initialize non-informative Bayesian exponential degradation model (Eq. 7.15, 7.16)

   (b) Hierarchical model updating with $S_{svm}(x_{new})$

   (c) Estimate time until fault develops using proposed prognostic t^3 (Eq. 7.17-7.18)

7.4 Validation Approach

Proper validation of the B-OCSVM approach requires complete run-to-failure data on which the proposed OC-SVM hyperparameter estimation, degradation thresholding, and t^3 fault prediction approaches can be applied. Furthermore, demonstrating the robustness and generalization ability of B-OCSVM requires validation to be performed across multiple datasets containing different fault types, sampling conditions and assets. Hence, validation is performed using three run-to-failure datasets: the IMS bearing dataset [101], the FEMTO bearing dataset [133] and the C-MAPSS turbofan engine dataset [163]. Each dataset serves the following purposes:

1. IMS Bearing Dataset: general validation of the methodology, benchmarking and sensitivity analysis of proposed OC-SVM hyperparameter estimation approach

2. FEMTO Bearing Dataset: evaluates the ability of B-OCSVM to generalize across different operating conditions, fault types, as well as bearings experiencing multiple types of faults simultaneously

3. C-MAPSS Turbofan Engine Dataset: evaluates the generalization ability of B-OCSVM when applied to applications outside of bearing faults (i.e. turbofan engine faults); and, assesses the performance of B-OCSVM when applied to sparsely sampled data

7.4.1 Evaluation Metrics

Due to the experimental conditions used to generate each of the validation datasets, ground truth labels for start of the degradation process and time of fault manifestation are not

available to calculate traditional performance metrics such as root mean squared error (RMSE). Hence, the performance of B-OCSVM will be quantified and compared to other approaches using the following proposed metrics:

1. First detection time (t_{FD}): time at which the degradation process is detected by B-OCSVM
2. Fault prediction time (t_{FP}): the time at which the degradation process is predicted to manifest into a fault based on the proposed degradation threshold $T_{S_{svm}}$
3. Earliest prediction time (t^3_{EP}): earliest time at which an accurate prediction of t_{FP} is obtained using t^3, which is quantified as the time the confidence interval for t^3 converges with the true RUL trajectory
4. Maximum actionable time (t_{FD-F}): the time between first detection and failure, which quantifies the maximum amount of time made available by B-OCSVM to take preventative action

For benchmarking, the goal is to compare the values of t_{FD} and t_{FP} obtained using B-OCSVM to the first fault detection times obtained by other approaches. Early detection would imply that t_{FD} in B-OCSVM precedes the first fault detection time of other approaches, while close alignment of t_{FP} to those fault detection times would indicate that the proposed degradation threshold provides a reasonable representation of the fault manifestation time. The remaining metrics t^3_{EP} and t^3_{EP} will be used to quantify the practical benefits (i.e. extra time to perform maintenance) provided by B-OCSVM.

7.5 Dataset Descriptions

This section describes each of the run-to-failure datasets used for validation of the proposed B-OCSVM approach.

7.5.1 IMS Run-to-Failure Bearing Dataset

The IMS dataset [101] provided by the Center for Intelligent Maintenance Systems (IMS) at the University of Cincinnati contains run-to-failure data for four bearings. The test rig consists of four force lubricated double row Rexnord ZA-2115 bearings mounted on

Test	# Samples	Faulty Bearing	Type	Fail. Time	Designation
S1	2156	B3	IRF	35 days	S1B3
S1	2156	B4	RF	35 days	S1B4
S2	984	B1	ORF	8 days	S2B1
S3	6322	B3	ORF	40 days	S3B3

Table 7.1: IMS dataset specifications [101]: IRF (inner race fault), RF (roller fault), ORF (outer race fault)

a shaft coupled to an AC motor rotating at 2000 RPM with a radial load of 6000 lbs. Dual-channel data collection was facilitated by two PCB 253B33 high sensitivity Quart ICP accelerometers mounted in the x and y directions. The sensing configuration and test bench layout are shown in Fig. 7.6 a and b, respectively. Data corresponding to four faulty bearing was collected over three independent run-to-failure tests. One second of vibration data is sampled at 20.48 kHz at 10 minute intervals until failure. Failure is defined as the time when the measured accumulated debris exceeds a specified threshold, which automatically terminates the test. Three independent run to failure tests were conducted, which resulted in a total of four failed bearings

Key Characteristics

The IMS dataset has been widely used for validation in machinery diagnostics, making it a suitable dataset for not only validation but also benchmaking. Table 7.1 summarizes the specifications for each failed bearing in the IMS dataset. The IMS bearing set contains three different types of bearing faults: inner race fault, roller fault and two outer race faults, with failure times ranging from 8 days to 40 days. Each test is conducted without any artificial defect initiation, resulting in more realistic degradation behaviour. The degradation behaviour varies significantly between each failed bearing, and this is reflected in the degree of difficulty for analysis. As a direct result, many methods have elected to analyze only a subset of the IMS bearings (shown in Table 7.5 hereafter). Successful detection and modeling of the fault mechanism across all four bearings requires a robust analysis approach. For validation and benchmarking, B-OCSVM is applied to all four faulty bearings in the dataset.

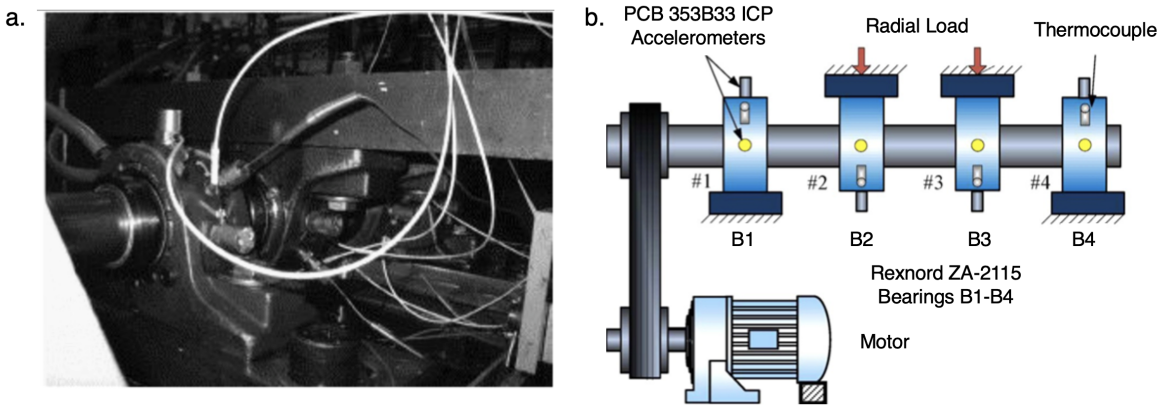

Figure 7.6: IMS Test Rig a. Sensing configuration b. Layout [58]

7.5.2 FEMTO Run-to-Failure Bearing Dataset

The FEMTO dataset is a run-to-failure bearing dataset provided for the IEEE PHM 2012 Prognostic Challenge. Run-to-failure bearing data was collected using the PRONOSTIA accelerated aging platform shown in Fig. 7.7. The PRONOSTIA platform uses a two shaft configuration containing an asynchronous 250W motor on the primary shaft connected via gearbox to the secondary shaft which houses the test bearing. The maximum speed of the motor is 2830 rpm, which corresponds to a maximum secondary shaft speed of 2000 rpm. A radial load up to 4000 N can be applied to the test bearing to simulate a variety of loading conditions. The dataset contains 17 run-to-failure data (6 complete, 11 truncated), collected using a dual-channel (x and y) accelerometer configuration, sampled at 25.6 kHz for a duration of 0.1s every 10s. Failure is defined as the time when the measured acceleration exceeds 20g. It is presumed that each failed bearing experiences contains almost all types of possible faults [133].

Key Characteristics

Similar to the IMS dataset, each test into the FEMTO dataset was initiated with no artificial defects, resulting in more realistic degradation behaviour. Contrary to the IMS dataset however, the failed bearings in the FEMTO dataset were obtained across several different operating conditions and are presumed to contains almost all types of possible faults. Additionally, the failed bearings exhibit a wide range of degradation behaviours, which adds further complexity to the analysis. These qualities make the FEMTO dataset

an excellent tool for demonstrating the ability of a method to generalize across different operating states and bearings containing multiple types of faults. The data specifications are provided in Table 7.2. to demonstrate generalization performance, B-OCSVM is applied to all six complete run-to-failure bearing data.

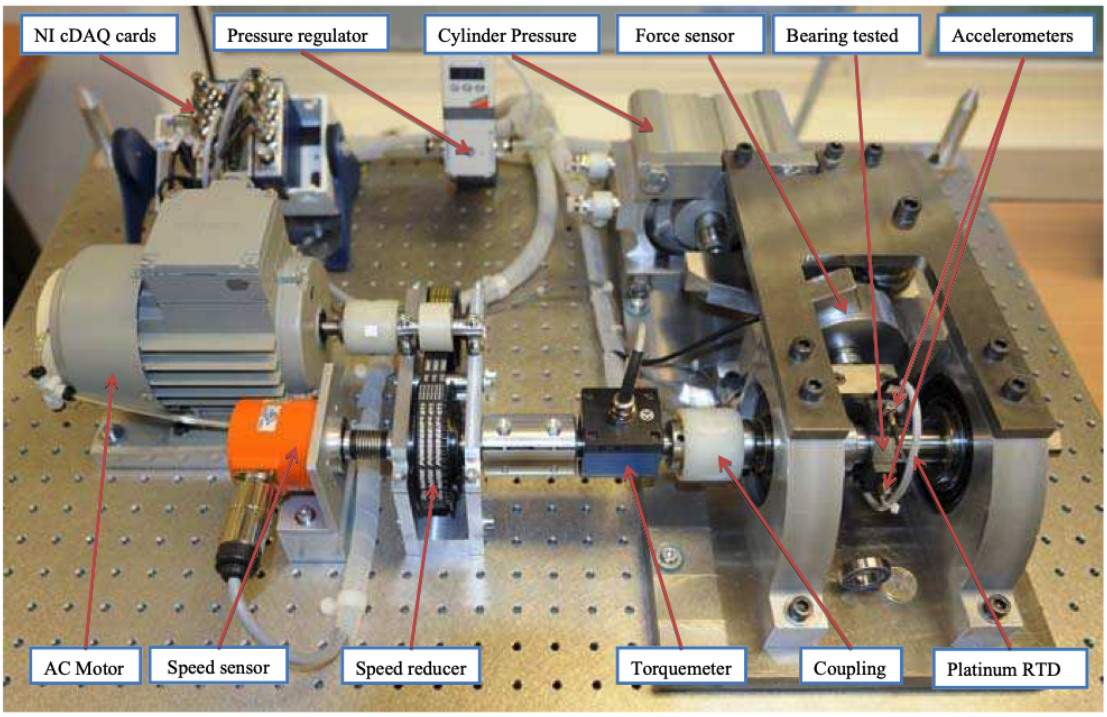

Figure 7.7: FEMTO Test Rig Layout [133]

7.5.3 C-MAPSS Run-to-Failure Turbofan Engine Dataset

The commercial modular aero-propulsion system simulation (C-MAPSS) dataset contains simulated run-to-failure turbofan engine data which are generated using the C-MAPSS simulation platform [163]. The engine simulated is of the 90,000 lb thrust class, with user configurable conditions including altitude, mach number, sea level temperature and thrust level. The built-in control system consists of a fan-speed controller, three high-limit regulators to prevent the engine from exceeding the designed core speed, engine-pressure ratio and high pressure turbine (HPT) exit temperature, and acceleration and deceleration limiters for the core speed. The configuration of the simulated engine is shown in Fig. 7.8.

Bearing	# Samples	Speed (rpm)	Load (N)	Lifetime (hrs)
B11	2803	4000	1800	7.8
B12	871	4000	1800	2.4
B21	911	4200	1730	2.5
B22	796	4200	1730	2.2
B31	515	5000	1500	1.4
B32	1636	5000	1500	4.5

Table 7.2: FEMTO dataset specifications [133]

The simulation dataset is divided into four subsets (FD001-FD004), outlined in Table 7.3, each containing different operating conditions and fault types. In every single test, the engine is initiated with a varying degree of initial wear and manufacturing variation, which introduces slight variations in degradation behaviour across machines, even within the subset. Multivariate sensor data is collected across 21 sensors placed at different locations in the engine.

Key Characteristics

Unlike the IMS and FEMTO bearing datasets, the C-MAPSS dataset provides run-to-failure data for a different type of mechanical asset - the turbofan engine, which provides several unique qualities for validation. Unlike the IMS and FEMTO bearing datasets, in which analysis is performed on a singular mechanical component, analysis on the C-MAPSS dataset can be performed at the system level (i.e. entire engine assembly). The other key difference is in the sampling rate - the C-MAPSS data is sampled at 1Hz, which is significantly lower than the sampling rate used in the IMS and FEMTO datasets, which poses a problem for many analysis approaches, particularly those heavily reliant upon frequency and time-frequency methods. Hence, the C-MAPSS dataset is an excellent choice for further demonstrating the ability of a method to generalize across different applications, and for evaluating performance in the case of sparsely sampled data. The proposed approach is applied to the first six complete run-to-failure data (denoted as Mach. 1-6) from the FD001 subset.

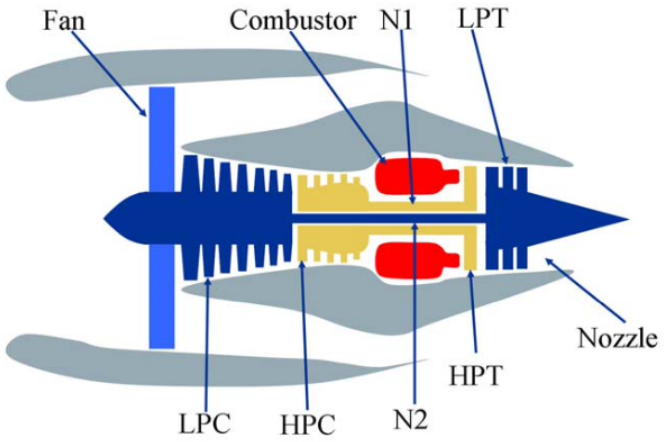

Figure 7.8: C-MAPSS simulated engine configuration [163]

Parameter	FD001	FD002	FD003	FD004
Engines for training	100	260	100	249
Engines for testing	100	259	100	248
Simulation conditions	1	6	1	6
Fault modes	1	1	2	2

Table 7.3: C-MAPPS dataset specifications []

7.6 Application of B-OCSVM to IMS Dataset

The B-OCSVM approach is applied to all four failure conditions in the IMS dataset. To maintain consistency for benchmarking, the test/train partitions for each bearing are based on those used in [110]. Feature extraction, dimensionality reduction, outlier removal using the proposed NN-based approach Eq. 7.7 is applied to the training partition of each bearing. Following outlier removal, ν is set to 0.01 [54]. Fig. 7.9 illustrates ΔD_k (Eq. 7.6), the outlier threshold (Eq. 7.7) and fraction of outliers η for each bearing. The optimal value of γ for each bearing is found by minimizing the objective functions (Eq. 7.8) shown in Fig. 7.10 and taking the central value (γ_{opt}) of the minimum region. Fig. 7.10 shows that a region where $J(\gamma)$ is at a minimum exists for each case.

The OC-SVM classifiers are trained using γ_{opt} and $\nu = 0.01$ and the scaled sigmoid-normalized OC-SVM distances (S_{svm}) for the test partitions shown in Fig. 7.11 are obtained

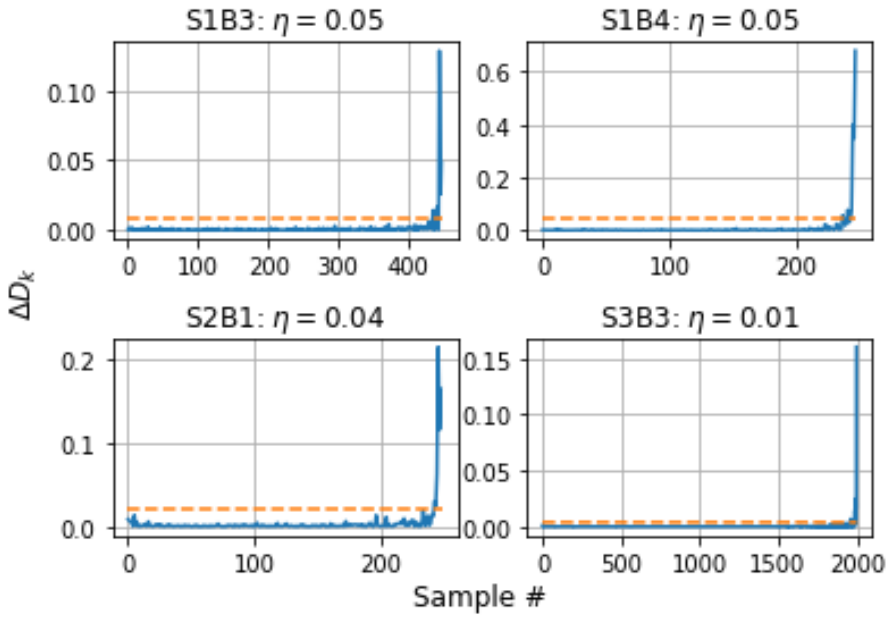

Figure 7.9: First difference of sorted k-NN distances ΔD_k and outlier threshold D_{cut} (dashed line) for each faulty bearing in IMS dataset

using Eq. 7.10. The early degradation process is detected using Eq. 7.14 to obtain t_{FD} for each bearing. At t_{FD}, a Bayesian degradation model is initialized with $\phi = S_{svm}(t_{FD})$, and non-informative priors $\theta \sim N(-1, 10^6)$, $\beta \sim N(1, 10^6)$. Updated model parameters are updated by computing the posteriors after each subsequent observation, and an estimate for t^3 is generated after each update until the earliest prediction time (t^3_{EP}) can be obtained. The degradation model for bearing S1B3 and corresponding probability density function of t^3 are shown in Fig. 7.12a and b, respectively. The metric t^3_{EP} is obtained from the $\alpha - \lambda$ plot (Fig. 7.13) as the earliest time at which the confidence interval for t^3 converges with the true RUL trajectory ($t^3_{EP} = 1580$ in Fig. 7.13) [162].

7.6.1 Results

The resulting B-OCSVM first detection times (t_{FD}), earliest prediction times (t^3_{EP}), fault prediction times (t_{FP}) and maximum actionable times (t_{FD-F}) on the IMS dataset are summarized in Table 7.4. In all cases, the degradation process is identified by t_{FD} well in advance of failure. Additionally, Fig. 7.11 shows that prior to t_{FD}, the behaviour of

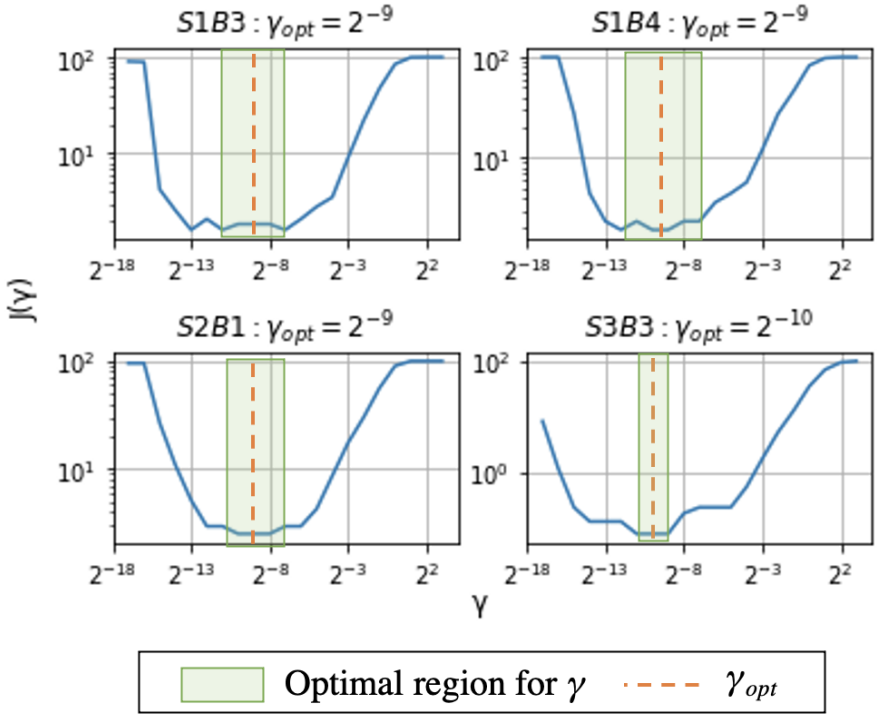

Figure 7.10: Objective function $J(\gamma)$, corresponding optimal region of γ and γ_{opt} for IMS dataset.

S_{svm} is relatively stable with little to no trend. Post-t_{FD} however, S_{svm} begins to deviate sharply from the healthy class, indicating that B-OCSVM correctly captures the start of the degradation process. In terms of the maximum actionable time, the values of t_{FD-F} in Table 7.4 show that B-OCSVM provides at least several days in each case to take preventative action before failure occurs.

7.6.2 Benchmarking

In order to validate the claims of early detection and fault prediction threshold, B-OCSVM is benchmarked against the current state-of-the-art supervised, semi-supervised and unsupervised fault detection approaches using the IMS dataset. Table 7.5 summarizes the performance of B-OCSVM against several comparable state-of-the-art approaches. Note that for the other approaches, t_{FD} in Table 7.5 represents the earliest time at which a fault is detected (i.e. first detection time). With the exception of S2B1, Table 7.5 shows that

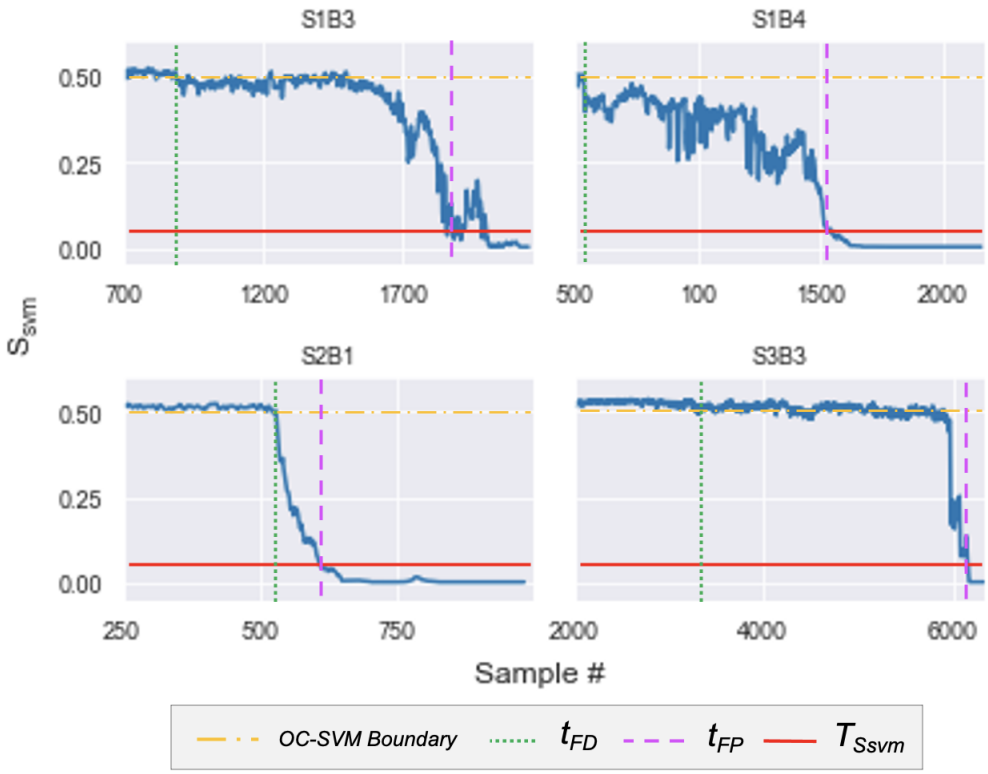

Figure 7.11: Scaled logistic sigmoid normalized OC-SVM distance (S_{svm}) on IMS Dataset

t_{FD} obtained using B-OCSVM occurs well in advance of those obtained using other methods. Comparing the B-OCSVM values of t_{FP} to the values of t_{FD} obtained using other approaches finds that there is close alignment in S1B3, S1B4 and S3B3, which demonstrates the that proposed threshold $T_{S_{svm}}$ accurately captures the level of degradation that would be typically considered a fault by other approaches. The exception to this is S2B1, in which t_{FD} occurs only slightly earlier in B-OCSVM and t_{FP} occurs slightly later than in most approaches. Also of significance is the fact that the detection sensitivity achieved using the unprocessed handcrafted feature set is comparable to those obtained using black boxing feature learning approaches.

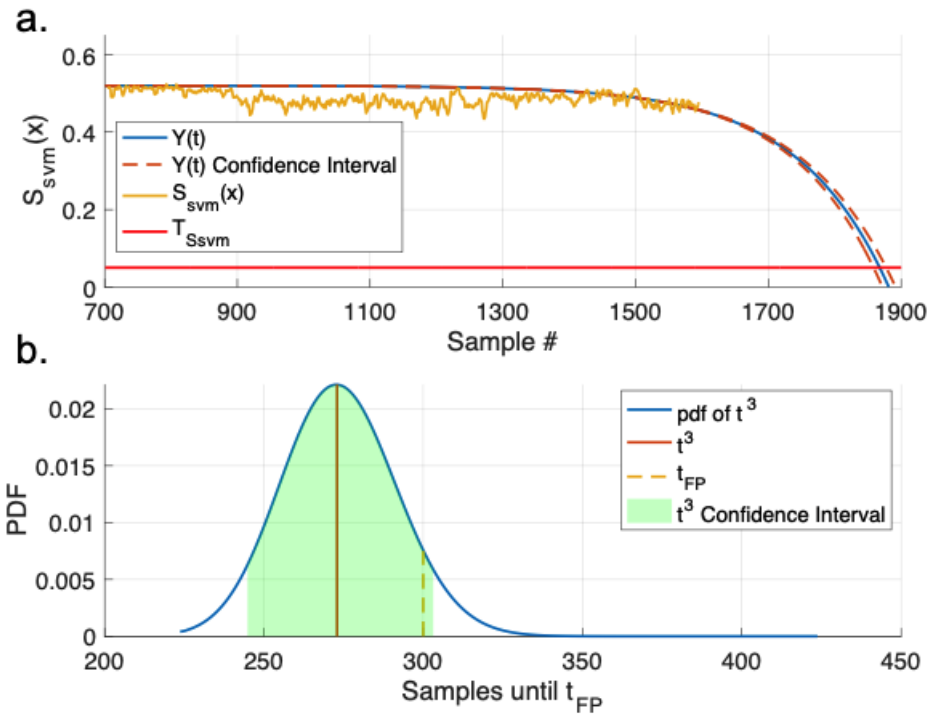

Figure 7.12: (a) Bayesian degradation model $Y(t)$ for IMS bearing S1B3 (t = 1580) (b) Corresponding PDF of t^3

7.6.3 Practical Significance of B-OCSVM and t^3 to Maintenance Planning

To illustrate the practical significance of the early-detection capabilities of B-OCSVM and prognostic t^3, several additional quantitative time-based metrics obtained using B-OCSVM are compared to those obtained using a comparable state of the art approach (CSC-NSVDD [110]). These results are shown in Table 7.6. With the exception of S3B3, where the degradation occurs very suddenly prior to failure, it can be shown by the metric $t_{FP} - t^3_{EP}$ that B-OCSVM and the t^3 prognostic provides a significant window of time to take preventative actions before a typical fault manifests. Furthermore, comparing the time until failure from first detection (t_{FD-F}) of B-OCSVM to those obtained using CSC-NSVDD, it was found that for all four cases B-OCSVM also offers a significantly longer window to take preventative actions before failure occurs.

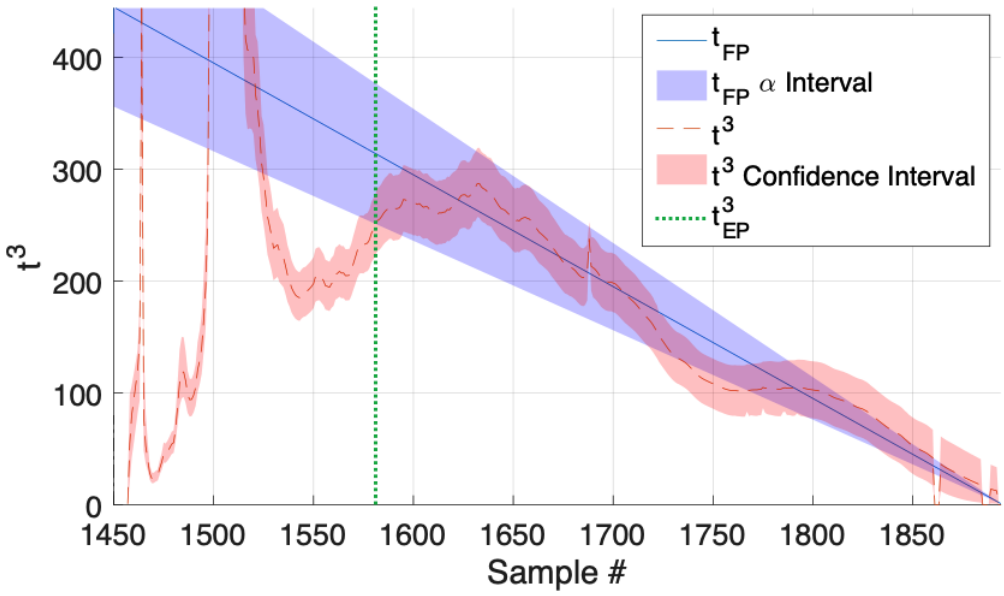

Figure 7.13: α-λ plot for IMS bearing S1B3

7.6.4 Sensitivity Analysis of OC-SVM Hyperparameter Approach

The performance of the proposed OC-SVM hyperparameter estimation approach is assessed by evaluating the influence of the estimated values of γ on the detection results. Specifically, the robustness of t_{FD} and t_{FP} to the choice of γ is demonstrated by computing t_{FD} and t_{FP} for the boundary values of γ shown in Fig. 7.10. The results are presented in Table 7.7. In Table 7.10, $\Delta - max$ represents the maximum deviation of the detection times with respect to those obtained using γ_{opt}, and demonstrates that t_{FD} remains unchanged for all values of γ with the optimal region. In regards to the t_{FP}, nominal changes are observed for S1B3 and S3B3, while a larger deviation of $\approx 12\%$ is observed in S1B4 and S2B1. This result is acceptable given that these deviations are associated with the extreme (boundary) values of γ and conclude that choice of γ, provided that is selected from within the optimal region, will not meaningfully influence the detection results.

Measure	S1B3	S1B4	S2B1	S3B3
t_{FD}	891	508	523	3289
t^3_{EP}	1580	1274	568	6010
t_{FP}	1844	1508	639	6071
Failure	2156	2156	984	6322
t_{FD-F} (days)	**4.0**	**6.1**	**2.9**	**2.2**

Table 7.4: First detection time (t_{FD}), earliest prediction time (t^3_{EP}) obtained using t^3, fault prediction time (t_{FP}) and maximum actionable time (t_{FD-F}) on IMS dataset using B-OCSVM

Approach	S1B3	S1B4	S2B1	S3B3
B-OCSVM	**891** 1844	**508** 1508	**523** 639	3289 6071
CSC-NSVDD [110]	1796	831	532	5973
AEC [67]	2027-2120	1641-1681	547-610	**2367-2435**
SODRMB-S4SVM [124]	-	-	535	-
SDAE-LSTM [169]	1812	1572	540	-
GAN [12]	1626-1684	1179-1192	-	-
MAS-Kurtosis [93]	1910	1650	710	No det.
DPCA-HMM [201]	2120	1760	539	-
VRCA [201]	-	1727	-	No det.
RMS [67]	2094	1730	539	No det.

Table 7.5: First detection times (t_{FD}) on IMS dataset based on literature from 2012 to 2020. For B-OCSVM, the first value for each case (t_{FD}) denotes the initial detection of the degradation process, while the second value denotes the fault prediction time (t_{FP}). '-' indicates when a bearing was not analyzed

7.7 Application of B-OCSVM to FEMTO Dataset

The purpose of the FEMTO dataset and subsequent C-MAPSS datasets are to provide supplemental validation of the generalization performance of B-OCSVM across additional scenarios. The FEMTO dataset specifically is used here to demonstrate the ability of B-OCSVM to generalize across different operating conditions, failure types, and bearings containing more than one type of fault. The proposed approach is applied to the six complete run-to-failure bearing data obtained across three discrete operating states as

Measure	S1B3	S1B4	S2B1	S3B3
$t_{FP} - t^3_{EP}$	1.8 days	1.7 days	0.9 days	0.4 days
t_{FD-F} (B-OCSVM)	**4 days**	**6.1 days**	**2.9 days**	**2.2 days**
t_{FD-F} (CSC-NSVDD)	2.2 days	4.6 days	1.7 days	0.3 days

Table 7.6: Time to fault from t^3_{EP} ($t_{FP} - t^3_{EP}$) and maximum actionable time (t_{FD-F}) on IMS dataset using B-OCSVM and CSC-NSVDD [110]

γ	S1B3		S1B4		S2B1		S3B3	
Lower Bound	891	1883	508	1581	523	639	3289	6158
Central Value	891	1844	508	1508	523	639	3289	6071
Upper Bound	891	1832	508	1323	523	561	3289	5968
Δ-max	0%	2.1%	0%	-12.3%	0%	-12.2%	0%	-1.7%

Table 7.7: Sensitivity of t_{FD} and t_{FP} on IMS dataset with respect to γ within the optimal region. Δ-max denotes the maximum change in detection point with respect to the central (optimal) value of γ

shown in Table. 7.2. The process used to apply B-OCSVM to the FEMTO dataset is identical to the approach used for the IMS dataset in section 7.6, with changes or adjustments made to the methodology between the two datasets.

7.7.1 Generalization to Different Operating States, Fault Types, Multi-Fault Bearings - Results

The degradation behaviour of S_{svm} and objective functions for each bearing into the FEMTO dataset are shown in Fig. 7.15 and 7.14, respectively. Fig. 7.15 shows that the FEMTO data contains a wide range of degradation paths. While the degradation behaviour captured by S_{svm} varies significantly between bearings, the proposed objective function ($J(\gamma)$) (Fig. 7.14 converges to an optimum region for all six bearings. The obtained results are summarized in Table 7.8. As was the case with the IMS dataset, the t_{FD}s obtained for each case occur well in advance of failure, and the unsupervised prediction for time of fault occurrence (t_{FP}) corresponding to the degradation threshold $T_{S_{svm}}$ precedes the actual point of failure in all cases. Furthermore, the degradation models with prognostic t^3 are able to successfully obtain a prediction of t_{FP} (t^3_{EP}) well in advance of fault occurrence in most cases except in operating state 3 (B31 and B32), in which the

fault manifests spontaneously. Lastly, relative to the total lifetime of each bearing found in Table 7.2, the maximum actionable times (t_{FD-F}) obtained using B-OCSVM provide a significant window to take preventative action before failure occurs. These results are consistent with those obtained using the IMS dataset and demonstrates that B-OCSVM generalizes well across different loading and speed conditions as well as on bearings containing multiple types of faults.

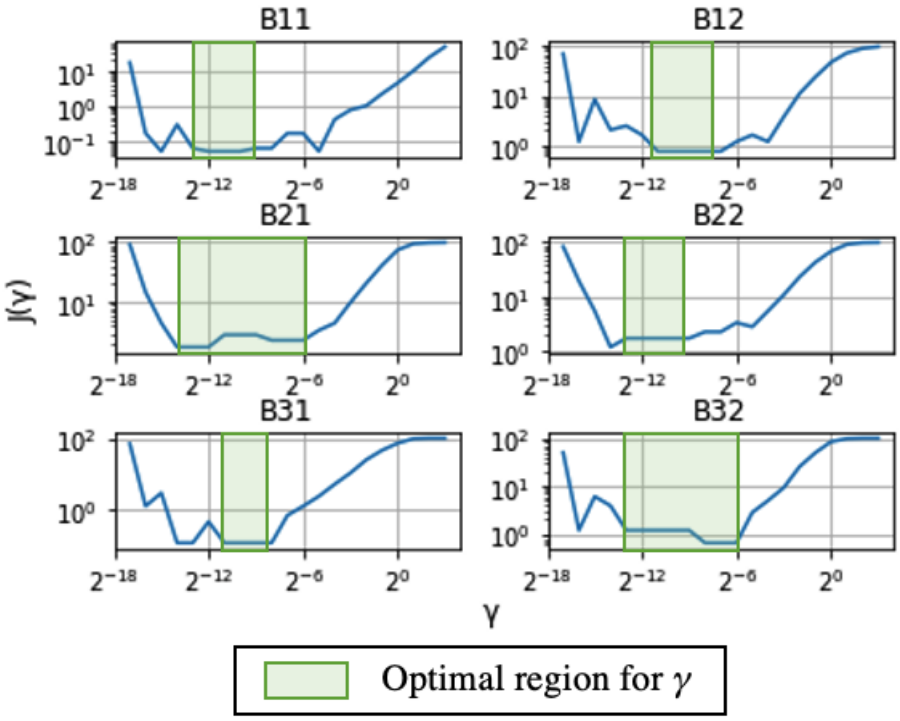

Figure 7.14: Objective function $J(\gamma)$, corresponding optimal region of γ for FEMTO bearing dataset

7.8 Application of B-OCSVM to C-MAPSS Turbofan Dataset

To demonstrate the ability of B-OCSVM to generalize to faults in other types of systems and to evaluate the performance of B-OCSVM with sparsely sampled sensor data in which

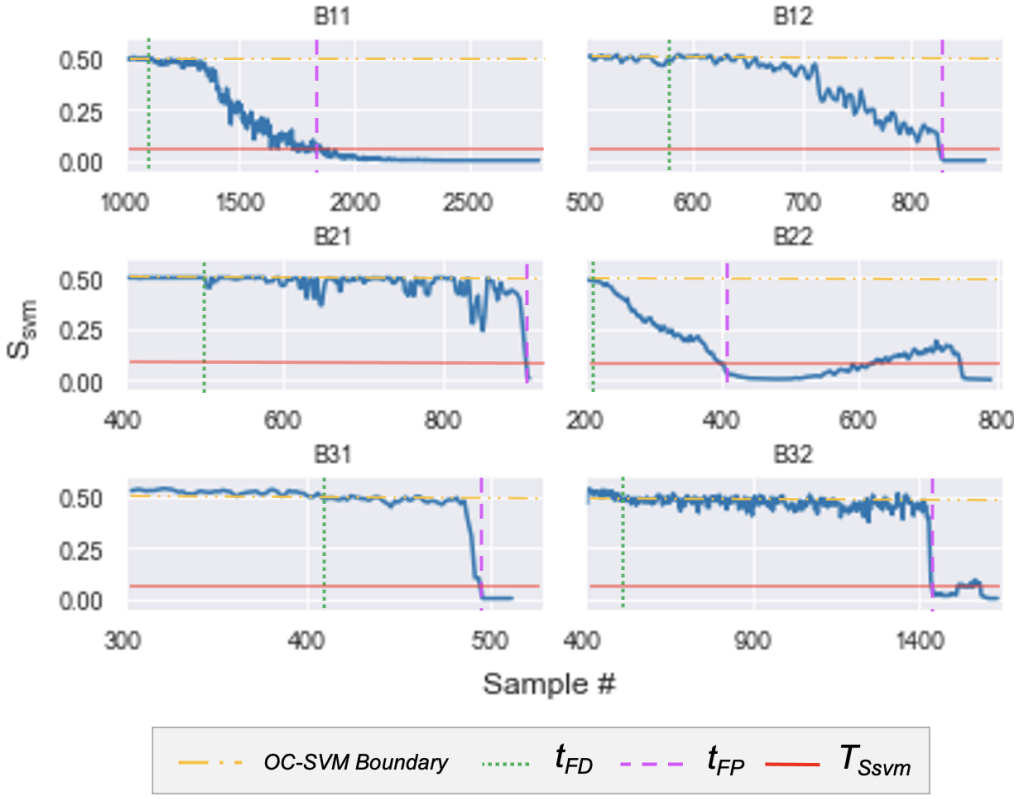

Figure 7.15: Scaled logistic sigmoid normalized OC-SVM distance (S_{svm}) on FEMTO Dataset

	Op. State 1		Op. State 2		Op. State 3	
Measure	B11	B12	B21	B22	B31	B32
t_{FD}	1128	564	498	204	408	503
t^3_{EP}	1414	664	753	225	478	1362
t_{FP}	1731	826	894	400	495	1434
Failure	2803	871	911	796	515	1636
t_{FD-F} (hrs)	**4.7**	**0.9**	**1.2**	**1.6**	**0.3**	**3.2**

Table 7.8: First detection time (t_{FD}), fault prediction time (t_{FP}), earliest prediction time (t^3_{EP}) and maximum actionable time (t_{FD-F}) on FEMTO bearing dataset using B-OCSVM

Measure	Mach. 1	Mach. 2	Mach. 3	Mach. 4	Mach. 5	Mach. 6
t_{FD}	76	130	60	86	99	71
t^3_{EP}	104	153	70	111	120	111
t_{FP}	150	223	128	141	201	154
Failure	190	285	177	187	267	186
t_{FD-F} (# cyc.)	**114**	**155**	**117**	**101**	**168**	**115**

Table 7.9: First detection time (t_{FD}), fault prediction time (t_{FP}), earliest prediction time (t^3_{EP}) and maximum actionable time (t_{FD-F}) on C-MAPSS turbofan dataset using B-OCSVM

the proposed feature set cannot be extracted, the proposed approach is applied to the first six run-to-failure data (Mach. 1-6) from the FD001 subset of the C-MAPSS Turbofan dataset. Since the sampling rate of the data is too low for proposed feature set, the raw data from all 21 sensors is used instead. Apart from the feature set used, the methodology used here is identical to approach applied to the previous two datasets.

7.8.1 Generalization to Turbofan Engine Faults and Sparse Data - Results

The objective functions ($J(\gamma)$) and degradation paths (S_{svm}) for the six FD001 turbofan data are shown in Fig. 7.16 and 7.17, respectively. Given that subset FD001 contains only one operational condition, the degradation behaviour is consistent across all six engines, with slight variations due to random initial defects. Similar to the IMS and FEMTO datasets, Fig. 7.16 shows that an optimal region for γ is obtained for each of the six engines. The obtained results are summarized in Table 7.9. From Table 7.9, the main observations regarding t_{DT}, t^3_{EP},t_{FP} and t_{FD-F} are consistent with those obtained on the IMS and FEMTO datasets, which indicates that B-OCSVM generalizes well to different faults in other machinery and performs well even in cases where the proposed feature set cannot be used.

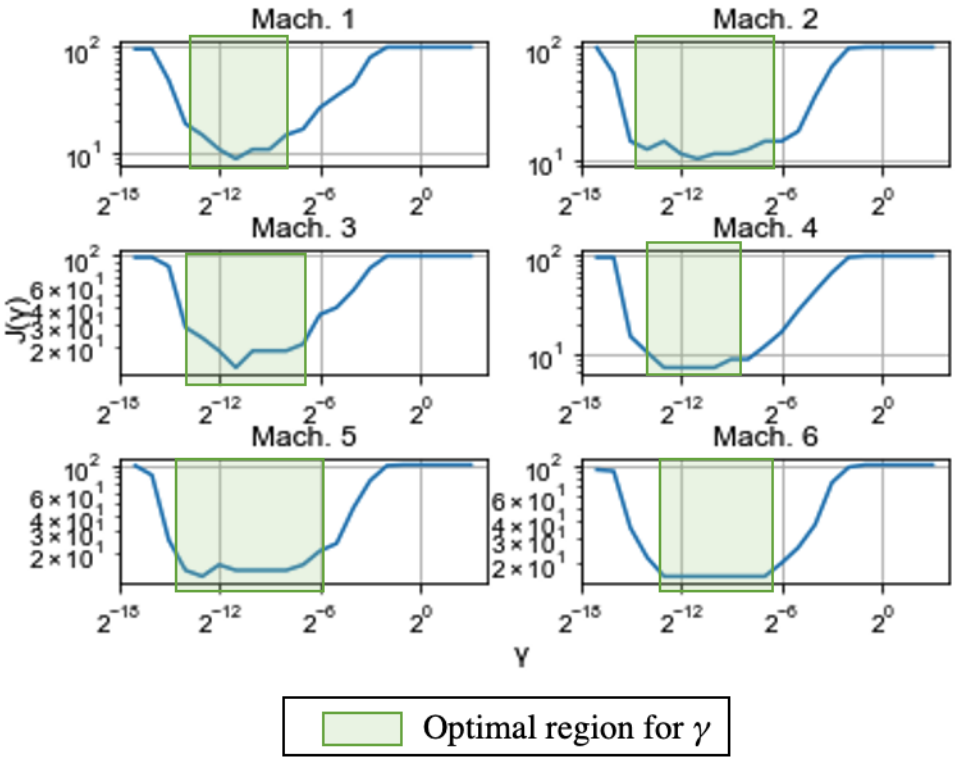

Figure 7.16: Objective function $J(\gamma)$, corresponding optimal region of γ for C-MAPPS dataset

7.9 Summary

This chapter presents an unsupervised framework for early degradation detection and fault prediction (B-OCSVM), based on OC-SVM and the hierarchical Bayesian framework. Within this approach, a new method for unsupervised OC-SVM hyperparameter estimation in the context of early degradation detection and a new prior-independent degradation threshold ($T_{S_{svm}}$) based on the OC-SVM decision space are proposed. he proposed threshold can be used in conjunction with a novel prognostic t^3 to predict the time at which the detected degradation process will manifest into a fault without prior knowledge regarding the failure of the component. Validation and benchmarking against the state-of-the-art was performed using multiple data sets. The key conclusions from this work are as follows:

1. Benchmarking against state-of-the-art fault detection approaches on the IMS data

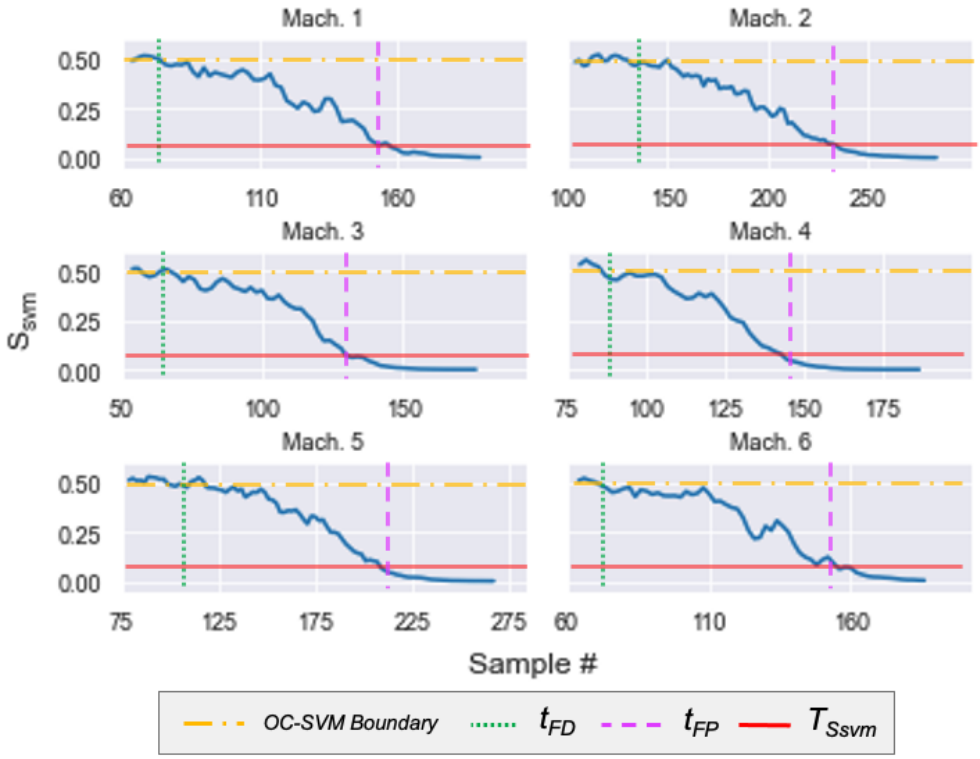

Figure 7.17: Scaled logistic sigmoid normalized OC-SVM distance (S_{svm}) on C-MAPPS dataset

set demonstrated that B-OCSVM was able to identify the start of the degradation process well in advance of the typical detection point found using other approaches, while still correctly predicting the fault manifest point identified by other approaches. These results demonstrate the early degradation detection capabilities of B-OCSVM, and validates that the proposed degradation threshold $T_{S_{svm}}$ is a good indicator of the fault manifestation point.

2. Sensitivity analysis of the proposed OC-OCSVM hyperparameter approach demonstrated that the choice of γ within the estimated optimal range does not meaningfully impact the detection results, which demonstrates the robustness of the parameter estimation approach.

3. Direct compared of the maximum actionable time t_{FD-F} obtained using B-OCSVM to CSC-NSVDD on the IMS dataset showed that B-OCSVM provides a significantly

longer window to take preventative action across all four failed bearings.

4. Validation performed on the FEMTO data set demonstrated the robustness of B-OCSVM to different operating states and bearings containing multiple types of faults.

5. Validation performed on the C-MAPSS turbofan engine data set demonstrated the ability of B-OCSVM to generalize to different machinery types, and showed that B-OCSVM can still be effective even in situations where the proposed feature set cannot be extracted.

Chapter 8

Contributions and Conclusions

In response to the unique challenges introduced by the increasing complexity of engineering systems—and the limited applicability of existing solutions within literature to these problems—this thesis presents a novel set of three, autonomous and unsupervised tools towards the realization of CBM on such systems. The proposed tools address several of the key knowledge gaps within the current body of unsupervised CBM literature pertaining to the areas of pre-processing, fault detection and degradation modeling. Validation of the proposed approaches is performed using a combination of data collected from two full-scale field implementations at Toronto Pearson International Airport, in addition to five, competition and public validation data sets. This chapter summarizes the significant contributions and key conclusions from this research, followed by several recommendations for future study.

8.1 Summary of Contributions

The main contributions of this thesis are as follows:

1. An autonomous, non-parametric pre-processing tool for spectral analysis of non-stationary signals (S-MSC) which uses MSC in conjunction with the STFT to blindly separate time-varying harmonic components from noise within a signal. The proposed S-MSC approach addresses the key gap in literature pertaining to the lack of computationally efficient, fully-blind, pre-processing methods conducive to unsupervised CBM. Validation performed across multiple use cases and data sets demonstrated

the robustness and effectiveness of S-MSC as an non-parametric pre-processing tool for non-stationary signals.

   (a) Collection of angular measurements from the LINK APM gearbox for validation

   (b) Data-driven derivation of a suitable estimator for the kernel bandwidth to enable autonomous deployment of S-MSC in the unsupervised setting.

2. An unsupervised pre-processing tool for blind decomposition of vibration signals obtained from multi-modal machinery (GMM-OSD), which uses GMMs to cluster and classify data within a non-stationary multi-modal vibration signal into a set of constituent, stationary signals. Applied in conjunction with S-MSC, the combined tools form a robust pre-processing framework for the efficient, unsupervised analysis of vibration signals containing multiple types of non-stationarity. The GMM-OSD approach and combined S-MSC GMM-OSD approach was validated using data obtained from two field pilots located at Toronto Pearson International Airport.

   (a) Collection and hand-labelling of multi-modal data sets from both the LINK APM gearbox and Terminal 1 PCA units; and, feature-based characterization of the LINK APM gearbox operational behaviour for validation

3. An unsupervised framework for early degradation detection and fault prediction (B-OCSVM), which builds upon OC-SVM and the hierarchical Bayesian framework. B-OCSVM addresses two key gaps in the literature pertaining to hyperparmeter estimation and degradation thresholding in the unsupervised setting. By re-framing the traditional fault detection and degradation modeling problem as a degradation detection and fault prediction problem, novel unsupervised approaches for both OC-SVM hyperparameter estimation and degradation threshold setting could be proposed. Validation and benchmarking against the state-of-the-art was performed using multiple data sets.

4. Design, implementation and collection of data across two full-scale field pilots located at Pearson Toronto International Airport: one pertaining to the gearbox of the LINK APM train; and, the other pertaining to three passenger boarding tunnel PCA units located in Terminal 1.

5. Peer-reviewed journal and conference publications corresponding to each of the aforementioned contributions, with more forth-coming. A list of publications that have resulted directly from this work is provided in Appendix A.

8.2 Key Conclusions

For the sake of completeness and ease of contextualization with the contributions listed above, the key conclusions from each of the three proposed approaches are aggregated in this section.

Pertaining to S-MSC, the key conclusions are:

1. Validation performed on the LINK APM gearbox data showed that S-MSC was effective at separating the time-varying gearbox harmonics from noise.

2. Comparison to pre-processing using TSA showed that S-MSC was able to extract more harmonic components than TSA, without requiring the angular signal needed for TSA.

3. Application to the Safran engine data set showed that the use of S-MSC for de-noising in envelope analysis enabled the detection of time-varying bearing fault frequencies, circumventing the need for alternative, computationally expensive solutions.

4. Further validation against FSST and SSA using the CMMNO'14 wind turbine data showed that S-MSC performs better than FSST and equal to SSA on the low-SNR CMMNO'14 data, at a much lower computational cost.

5. The S-MSC approach depends upon only a single parameter, the kernel bandwidth. To enable autonomous deployment of S-MSC, a criterion based on the spectral skewness of the S-MSC residual spectrum was proposed for the estimation of the kernel bandwidth.

Pertaining to GMM-OSD, the key conclusions are:

1. A qualitative study was performed on data obtained from the LINK APM gearbox and Terminal 1 PCA units to characterize the dynamic operational behaviour of these machines and and validate the proposed feature set. The results of the study showed that the best differentiating features varied between the two data sets, which justifies the use of a wider feature set in the unsupervised setting.

2. Applied to data obtained the LINK APM gearbox and Terminal 1 PCA unit field pilots, GMM-OSD was able to blindly decompose the multi-modal data in each signal into the corresponding discrete operating states. For the well-separated Terminal 1 PCA data, GMM-OSD could be applied effectively without pre-processing.

3. For the LINK APM gearbox data, GMM-OSD was unable to cluster the data effectively without pre-processing. After applying S-MSC first as a pre-processing tool to remove non-multi-modal sources of non-stationarity from the signal, GMM-OSD was able to decompose the filtered signal into the correct, discrete operating states.

4. A second formulation using vGMMs was also validated using both data sets, and was shown to converge to the same optimal model selected using BIC in the traditional approach. Both approaches were equally effective for decomposition, with vGMM providing a more computationally efficient solution, and GMM-BIC providing higher interpretability of the model choice due to the use of BIC.

Lastly, pertaining to B-OCSVM, the key conclusions are:

1. Benchmarking against state-of-the-art fault detection approaches on the IMS data set demonstrated that B-OCSVM was able to identify the start of the degradation process well in advance of the typical detection point found using other approaches, while still correctly predicting the fault manifest point identified by other approaches. These results demonstrate the early degradation detection capabilities of B-OCSVM, and validates that the proposed degradation threshold $T_{S_{svm}}$ is a good indicator of the fault manifestation point.

2. Sensitivity analysis of the proposed OC-OCSVM hyperparameter approach demonstrated that the choice of γ within the estimated optimal range does not meaningfully impact the detection results, which demonstrates the robustness of the parameter estimation approach.

3. Direct compared of the maximum actionable time t_{FD-F} obtained using B-OCSVM to CSC-NSVDD on the IMS dataset showed that B-OCSVM provides a significantly longer window to take preventative action across all four failed bearings.

4. Validation performed on the FEMTO data set demonstrated the robustness of B-OCSVM to different operating states and bearings containing multiple types of faults.

5. Validation performed on the C-MAPSS turbofan engine data set demonstrated the ability of B-OCSVM to generalize to different machinery types, and showed that B-OCSVM can still be effective even in situations where the proposed feature set cannot be extracted.

8.3 Limitations

The unsupervised methods presented in this thesis have been developed within the context of machinery monitoring. While unsupervised (i.e. no requirements for labelled data) inductive biases or assumptions drawn from domain knowledge are incorporated in the formulation of these methods. As such, while these methods are designed to generalize well within the realm of machinery monitoring, the generalization performance and applicability of these methods - without modification - to other applications outside of machinery monitoring, has yet to be evaluated.

8.4 Recommendations for Future Studies

While the tools proposed within this thesis provide a step forward towards a more effective implementation of CBM in the unsupervised setting, there remain several important topics that require further exploration:

1. **IoT Integration:** application of the proposed tools within the real-world will require further studies which focus specifically on their implementation within IoT. Namely, the feasibility of these tools over stream processing, and their integration into existing cloud computing services needs to be explored in greater detail.

2. **Multi-Component or Fleet-level Application:** all of the validation performed in this thesis considers only a single component or system at a time. Hence, two sensible extensions of this work are to explore their application within the context of multi-component and fleet-level CBM applications.

3. **Ensemble Learning:** To further improve the robustness of the proposed tools (namely B-OCSVM and GMM-OSD), particularly for challenging or highly uncertain real-world applications,extensions of this work to incorporate ensemble learning should be explored.

4. **Extension of CBM tools to Other Applications:** Relative to many other fields, the literature within CBM pertaining to DSP and ML is dense but pragmatic. Many existing methods, including those presented in this thesis, can be extended and explored in the context of other applications and other fields of engineering. The application of S-MSC to hydrophone signals as a spectral pre-processing tool for active leak detection by the author is only one such example [89]. A summary of this work is provided in Appendix B.

5. **Continued Exploration of Unsupervised CBM:** Relative to supervised CBM, the depth of literature within unsupervised CBM is comparatively shallow, leaving much left to be explored. While challenging to develop and implement, the practical implications of unsupervised CBM tools are highly significant and worthwhile to continue exploring.

References

[1] A. Aherwar and S. Khalid. Vibration analysis techniques for gearbox diagnostic: A review. *Int. J. Advanced Eng. Technol.*, 3(2):4–12, 2012.

[2] A. Ahrabian and D. P. Mandic. Selective Time-Frequency Reassignment Based on Synchrosqueezing. *IEEE Signal Process. Lett.*, 22(11):2039–2043, 2015.

[3] S. Ambani, L. Li, and J. Ni. Condition-based maintenance decision-making for multiple machine systems. *Journal of Manufacturing Science and Engineering-transactions of The Asme*, 131, 06 2009.

[4] M. Amer, M. Goldstein, and S. Abdennadher. Enhancing one-class support vector machines for unsupervised anomaly detection. In *Proc. ACM SIGKDD Workshop Outlier Detection Description*, page 8–15, 2013.

[5] A. Anaissi, K. Nguyen, and Y. Wang. Automated parameter tuning in one-class support vector machine: an application for damage detection. *International Journal of Data Science and Analytics*, 6, 12 2018.

[6] J. Antoni. The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20(2):282–307, 2006.

[7] J. Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mech. Syst. Signal Process.*, 21(1):108–124, 2007.

[8] J. Antoni et al. Feedback on the Surveillance 8 challenge: Vibration-based diagnosis of a Safran aircraft engine. *Mech. Syst. Signal Process.*, 97:112–144, 2017.

[9] I. Antoniadis and G. Glossiotis. Cyclostationary analysis of rolling-element bearing vibration signals. *Journal of Sound and Vibration*, 248(5):829–845, 2001.

[10] A. Ashasi-Sorkhabi, S. Fong, G. Prakash, and S. Narasimhan. A condition based maintenance implementation for an automated people mover gearbox. *International Journal of Prognostics and Health Management*, 8(020):13, 2017.

[11] B. Assaad, M. Eltabach, and J. Antoni. Vibration based condition monitoring of a multistage epicyclic gearbox in lifting cranes. *Mechanical Systems and Signal Processing*, 42(1-2):351–367, 2014.

[12] S. Baggeröhr, W. Booyse, P. S. Heyns, and D. N. Wilke. Novel bearing fault detection using generative adversarial networks. *Condition Monitoring Diagnostic Eng. Manag.*, page 243, 2018.

[13] K. Barbé and W. Van Moer. Automatic detection, estimation, and validation of harmonic components in measured power spectra: All-in-one approach. *IEEE Trans. Instrum. Meas.*, 60(3):1061–1069, 2011.

[14] R. G. Bartheld, T. G. Habetler, and F. Kamran. Motor Bearing Damage Detection Using Stator Current Monitoring. *IEEE Transactions on Industry Applications*, 31(6):1274–1279, 1995.

[15] E. Bechhoefer and M. Kingsley. A review of time synchronous average algorithms. *Annual Conf. Prognostics and Health Manage. Soc.*, pages 1–10, 2009.

[16] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak. Fault prognostic of bearings by using support vector data description. In *2012 IEEE Conference on Prognostics and Health Management*, pages 1–7, 2012.

[17] C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[18] A. Bousdekis, K. Lepenioti, D. Apostolou, and G. Mentzas. Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. *IFAC-PapersOnLine*, 52(13):607–612, 2019.

[19] L. Bull, K. Worden, G. Manson, and N. Dervilis. Active learning for semi-supervised structural health monitoring. *Journal of Sound and Vibration*, 437:373–388, 2018.

[20] S. Butterworth. On the Theory of Filter Amplifiers. *Experimental Wireless & the Wireless Engineer*, 7:536–541, Oct. 1930.

[21] W. Caesarendra, A. Widodo, and B.-S. Yang. Application of relevance vector machine and logistic regression for machine degradation assessment. *Mechanical Systems and Signal Processing*, 24(4):1161–1171, 2010.

[22] Y. Cao, N. Jan, B. Huang, M. Fang, Y. Wang, and W. Gui. Multimodal process monitoring based on variational bayesian pca and kullback-leibler divergence between mixture models. *Chemometrics and Intelligent Laboratory Systems*, 210:104230, 01 2021.

[23] C. Capdessus, M. Sidahmed, and J.-L. Lacoume. Cyclostationary processes: Application in gear faults early diagnosis. *Mechanical Systems and Signal Processing*, 14:371–385, 05 2000.

[24] Z. Chen and Z. Li. Fault diagnosis method of rotating machinery based on stacked denoising autoencoder. *Journal of Intelligent Fuzzy Systems*, 34:3443–3449, 2018.

[25] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8):790 –799, 1995.

[26] Y. Cheng et al. Application of an improved minimum entropy deconvolution method for railway rolling element bearing fault diagnosis. *J. Sound Vibration*, 425:53–69, 2018.

[27] Y. Cheng, C. Lu, T. Li, and L. Tao. Residual lifetime prediction for lithium-ion battery based on functional principal component analysis and Bayesian approach. *Energy*, 90:1983–1993, 2015.

[28] R. B. Chinnam, R. Babu, and P. Mohan. Online reliability estimation of physical systems using neural networks and wavelets. *International Journal of Smart Engineering System Design*, 4:253–264, 10 2002.

[29] CMMNO 14'. CMMNO'14 diagnosis contest. In *Int. Conf. Condition Monitoring Mach. in Non-Stationary Operation*, page 1, Lyon, 2014.

[30] M. Coats and R. Randall. Order-Tracking with and without a tacho signal for gear fault diagnostics. *Australian Acoustical Society Conference 2012, Acoustics 2012: Acoustics, Development, and the Environment*, (November):447–454, 2012.

[31] D. Comaniciu and P. Meer. Distribution free decomposition of multivariate data. *Pattern Anal. Appl.*, 2(1):22–30, 01 2000.

[32] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603 – 619, 2002.

[33] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[34] Z. Daher, E. Sekko, J. Antoni, C. Capdessus, and L. Allam. Estimation of the synchronous average under varying rotating speed condition for vibration monitoring. *ISMA2010 International conference on noise and vibration engineering*, (November 2014):2779–2788, 2010.

[35] X. Dai and Z. Gao. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, 9(4):2226–2238, 2013.

[36] H. J. Decker. Gear Crack Detection Using Tooth Analysis. *NASA TM 211491*, 2002.

[37] S. Depaoli. Mixture class recovery in gmm under varying degrees of class separation: frequentist versus bayesian estimation. *Psychological methods*, 18 2:186–219, 2013.

[38] F. Di Maio, K. Tsui, and E. Zio. Combining Relevance Vector Machines and exponential regression for bearing residual life estimation. *Mechanical Systems and Signal Processing*, 31:405–427, 2012.

[39] A. Diez-Olivan, J. A. Pagan, N. L. D. Khoa, R. Sanz, and B. Sierra. Kernel-based support vector machines for automated health status assessment in monitoring sensor data. *Int. J. of Advanced Manufacturing Tech.*, 95(1-4):327–340, 2018.

[40] P. Do, A. Voisin, E. Levrat, and B. Iung. A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. *Reliability Engineering and System Safety*, 133:22–32, 2015.

[41] S. Ebersbach, Z. Peng, and N. J. Kessissoglou. The investigation of the condition and faults of a spur gearbox using vibration and wear debris analysis techniques. *Wear*, 260(1-2):16–24, 2006.

[42] A. L. Ellefsen, E. Bjørlykhaug, V. søy, S. Ushakov, and H. Zhang. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliab. Eng. Syst. Saf.*, 183:240–251, 2019.

[43] A. H. Elwany, N. Z. Gebraeel, and L. M. Maillart. Structured Replacement Policies for Components with Complex Degradation Processes and Dedicated Sensors. *Operations Research*, 59(3):684–695, 2011.

[44] D. Fernández-Francos, D. Marténez-Rego, O. Fontenla-Romero, and A. Alonso-Betanzos. Automatic bearing fault diagnosis based on one-class m-SVM. *Computers and Industrial Engineering*, 64(1):357–365, 2013.

[45] S. Fong, A. Ashasi-Sorkhabi, G. Prakash, S. Narasimhan, and M. Riseborough. Automated condition-based monitoring of automated people movers. In *Proceedings of the 16th International Conference on Automated People Movers and Automated Transit Systems*, Tampa, FL, 2018. ASCE.

[46] S. Fong, J. Harmouche, S. Narasimhan, and J. Antoni. Mean shift clustering-based analysis of non-stationary vibration signals for machinery diagnostics. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4056–4066, 2020.

[47] S. Fong and S. Narasimhan. An unsupervised bayesian oc-svm approach for early degradation detection, thresholding, and fault prediction in machinery monitoring. *IEEE Transactions on Instrumentation and Measurement*, 71:1–11, 2022.

[48] S. Fong, S. Narasimhan, and M. Riseborough. A blind condition-based maintenance framework for real-time fault detection and degradation modeling of the link apm gearbox. In *Automated People Movers and Automated Transit Systems 2020*, pages 43–54. 2020.

[49] P. M. Frank and X. Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control*, 7(6):403–424, 1997.

[50] K. P. F.R.S. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[51] P. Gangsar and R. Tiwari. Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. *Mechanical Systems and Signal Processing*, 144:106908, 2020.

[52] N. Gebraeel, A. Elwany, and J. Pan. Residual life predictions in the absence of prior degradation knowledge. *IEEE Trans. Reliab.*, 58(1):106–117, 2009.

[53] N. Gebraeel and M. Lawley. A neural network degradation model for computing and updating residual life distributions. *Automation Science and Engineering, IEEE Transactions on*, 5:154 – 163, 02 2008.

[54] Z. Ghafoori, S. M. Erfani, S. Rajasegarar, J. C. Bezdek, S. Karunasekera, and C. Leckie. Efficient Unsupervised Parameter Estimation for One-Class Support Vector Machines. *IEEE Trans. Neural Netw. and Learn. Syst.*, 29(10):5057–5070, 2018.

[55] S. Ghassempour, F. Girosi, and A. Maeder. Clustering multivariate time series using Hidden Markov Models. *International journal of environmental research and public health*, 11(3):2741–2763, mar 2014.

[56] M. Goldstein and S. Uchida. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. 11(4):1–31, 04 2016.

[57] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[58] W. Gousseau, F. Girardin, and J. Griffaton. Analysis of the rolling element bearing data set of the center for intelligent maintenance systems of the university of cincinnati. 10 2016.

[59] D. Goyal and B. S. Pabla. Condition based maintenance of machine tools-A review. *CIRP J. Manufacturing Sci. Technol.*, 10:24–35, 2015.

[60] R. M. Gray and L. D. Davisson. *An introduction to statistical signal processing.* Cambridge University Press, 2004.

[61] J. Guo, Z. Li, and M. Li. A Review on Prognostics Methods for Engineering Systems. *IEEE Trans. Reliab.*, 69(3):1110–1129, 2020.

[62] K. Guo, D. Liu, Y. Peng, and X. Peng. Data-driven anomaly detection using ocsvm with boundary optimzation. In *2018 Conf. Prognostics Syst. Health Manag.*, pages 244–248, 2018.

[63] X. Guo, C. Shen, and L. Chen. Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery. *Applied Sciences*, 7(1):41, 2017.

[64] J. M. Ha, B. D. Youn, H. Oh, B. Han, Y. Jung, and J. Park. Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines. *Mechanical Systems and Signal Processing*, 70:161–175, 2016.

[65] J. Harmouche et al. The sliding singular spectrum analysis: A data-driven nonstationary signal decomposition tool. *IEEE Trans. Signal Process.*, 66(1):251–263, Jan 2018.

[66] T. J. Harvey, R. J. K. Wood, and H. E. G. Powrie. Electrostatic wear monitoring of rolling element bearings. *Wear*, 263(7-12 SPEC. ISS.):1492–1501, 2007.

[67] R. Hasani, G. Wang, and R. Grosu. A machine learning suite for machine components' health-monitoring. pages 9472–9477, 2019.

[68] V. Hautamaki, I. Karkkainen, and P. Franti. Outlier detection using k-nearest neighbour graph. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, pages 430–433 Vol.3, 2004.

[69] J. He, S. Yang, and C. Gan. Unsupervised Fault Diagnosis of a Gear Transmission Chain Using a Deep Belief Network. *Sensors*, 17(7), 2017.

[70] Q. He, X. Wang, and Q. Zhou. Vibration Sensor Data Denoising Using a Time-Frequency Manifold for Machinery Fault Diagnosis. pages 382–402, 2014.

[71] Y. He, C. Gu, Z. Chen, and X. Han. Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. *International Journal of Production Research*, 55(19):5841–5862, 2017.

[72] D. Hendrycks, M. Mazeika, and T. Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations*, 2019.

[73] R. Heng and M. Nor. Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition. *Applied Acoustics*, 53(1-3):211–226, 1998.

[74] P. S. Heyns, R. Vinson, and T. Heyns. Rotating machine diagnosis using smart feature selection under non-stationary operating conditions. *Insight - Non-Destructive Testing and Condition Monitoring*, 58(8):417–422, 2016.

[75] T. Heyns, P. S. Heyns, and J. P. De Villiers. Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. *Mechanical Systems and Signal Processing*, 32:200–215, 2012.

[76] G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, 2006.

[77] E. Hiroaki and S. Nader. Gearbox Simulation Models with Gear and Bearing Faults. *Mechanical Engineering*, pages 17–54, 2012.

[78] S. Ho, W. Chan, and H. Leung. Application of statistical signal processing for condition monitoring of rotor faults in induction motor. In *1993 Sixth International Conference on Electrical Machines and Drives (Conf. Publ. No. 376)*, pages 97–102, 1993.

[79] C.-W. Hsu and C.-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on neural networks*, 13(2):415–425, 2002.

[80] H. Z. Huang, H. K. Wang, Y. F. Li, L. Zhang, and Z. Liu. Support vector machine based estimation of remaining useful life: current research status and future trends. *Journal of Mechanical Science and Technology*, 29(1):151–163, 2015.

[81] J. Hulst. Modeling physiological processes with dynamic bayesian networks. *Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Pittsburgh*, 2006.

[82] R. Isermann. Process fault detection based on modeling and estimation methods—A survey. *Automatica*, 20(4):387–404, 1984.

[83] M. M. M. Islam and J.-M. Kim. Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network. *Computers in Industry*, 106:142–153, 2019.

[84] M. A. A. Ismail, N. Sawalhi, and A. Bierig. Fault Diagnosis of Wind Turbine Gearboxes using Enhanced Tacholess Order Tracking. *Ijmpe*, pages pp. 28–32, Volume–6,Issue–3, 2018.

[85] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke. Deep learning for infrared thermal image based machine health monitoring. *IEEE/ASME Transactions on Mechatronics*, 23(1):151–159, 2018.

[86] A. K. S. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

[87] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72-73:303–315, 2016.

[88] Q. Jiang and X. Yan. Multimode process monitoring using variational bayesian inference and canonical correlation analysis. *IEEE Transactions on Automation Science and Engineering*, 16(4):1814–1824, 2019.

[89] M. D. Kafle, S. Fong, and S. Narasimhan. Active acoustic leak detection and localization in a plastic pipe using time delay estimation. *Applied Acoustics*, 187:108482, 2022.

[90] X. Ke and Z. Xu. A model for degradation prediction with change point based on Wiener process. In *IEEE International Conference on Industrial Engineering and Engineering Management*, volume 2016-Janua, pages 986–990, 2016.

[91] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, Oct 2012.

[92] E. Y. Kim, A. C. C. Tan, J. Mathew, and B. S. Yang. Condition monitoring of low speed bearings: A comparative study of the ultrasound technique versus vibration measurements. *Australian Journal of Mechanical Engineering*, 5(2):177–189, 2008.

[93] S. Kim, S. Park, J. Kim, J. Han, D. An, N. H. Kim, and J. Choi. A new prognostics approach for bearing based on entropy decrease and comparison with existing methods. In *Annu. Conf. Prognostics Health Manag. Soc.*, 2016.

[94] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[95] I. Komorska and A. Puchalski. Rotating Machinery Diagnosing in Non-Stationary Conditions with Empirical Mode Decomposition-Based Wavelet Leaders Multifractal Spectra. *Sensors (Basel, Switzerland)*, 21(22):7677, nov 2021.

[96] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017.

[97] T. Le, M. Luo, J. Zhou, and H. L. Chan. Predictive maintenance decision using statistical linear regression and kernel methods. In *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, pages 1–6, 2014.

[98] M. Lebold, K. Mcclintic, R. Campbell, C. Byington, and K. Maynard. Review of vibration analysis methods for gearbox diagnostics and prognostics. In *54th Meeting of the Society for Machinery Failure Prevention Technology*, number November, pages 623–634, 2000.

[99] Q. Leclère, H. André, and J. Antoni. A multi-order probabilistic approach for Instantaneous Angular Speed tracking debrie fi ng of the CMMNO ' 14 diagnosis contest. *Mechanical Systems and Signal Processing*, 81:375–386, 2016.

[100] J. Lee, M. Ghaffari, and S. Elmeligy. Self-maintenance and engineering immune systems: Towards smarter machines and manufacturing systems. *Annual Reviews in Control*, 35:111–122, 04 2011.

[101] J. Lee, H. Qiu, G. Yu, J. Lin, and R. T. Services. Bearing data set. IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, 2007.

[102] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. *Mech. Syst. and Signal Process.*, 2014.

[103] Y. Lei, Z. He, Y. Zi, and Q. Hu. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mechanical Systems and Signal Processing*, 21(5):2280–2294, 2007.

[104] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. and Signal Process.*, 138:106587, 2020.

[105] C. J. Li and J. D. Limmer. Model-based condition index for tracking gear wear and fatigue damage. *Wear*, 241(1):26–32, 2000.

[106] D. Li, W. Wang, and F. Ismail. A fuzzy-filtered grey network technique for system state forecasting. *Soft Comput.*, 19(12):3497–3505, dec 2015.

[107] Y. Li and L. Maguire. Selecting critical patterns based on local geometrical and statistical information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(6):1189–1201, 2011.

[108] Y. Li, Y. Yang, X. Wang, B. Liu, and X. Liang. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine. *Journal of Sound and Vibration*, 428:72–86, 2018.

[109] J. J. Lin, J. Pulido, and M. Asplund. Reliability analysis for preventive maintenance based on classical and bayesian semi-parametric degradation approaches using locomotive wheel-sets as a case study. *Reliability Engineering [?] System Safety*, 134:143–156, 02 2015.

[110] C. Liu and K. Gryllias. A semi-supervised support vector data description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mech. Syst. and Signal Process.*, 140, 2020.

[111] D. Liu, J. Zhou, D. Pan, Y. Peng, and X. Peng. Lithium-ion battery remaining useful life estimation with an optimized relevance vector machine algorithm with incremental learning. *Measurement*, 63(Complete):143–151, 2015.

[112] J. Liu, A. Saxena, K. Goebel, B. Saha, and W. Wang. An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries. *Annual Conference of the Prognostics and Health Management Society, PHM 2010*, 01 2010.

[113] T. Liu, Z. Luo, J. Huang, and S. Yan. A comparative study of four kinds of adaptive decomposition algorithms and their applications. *Sensors*, 18:2120, 07 2018.

[114] W. Liu, G. Hua, and J. R. Smith. Unsupervised one-class learning for automatic outlier removal. In *2014 IEEE Conf. Comp. Vision Pattern Recognition*, pages 3826–3833, 2014.

[115] F. Locatello, S. Bauer, M. Lučić, G. Rätsch, S. Gelly, B. Schölkopf, and O. F. Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *Int. Conf. Mach. Learn.*, 2019.

[116] A. Lorenzoni, M. Kempf, and O. Mannuß. Degradation model constructed with the aid of dynamic Bayesian networks. *Cogent Engineering*, 4(1):1395786, 2017.

[117] M. Lou and T. Huang. An orthogonal technique for empirical mode decomposition in hilbert-huang transform. *MATEC Web of Conferences*, 24:03004, 01 2015.

[118] T. Loutas, D. Roulias, and G. Georgoulas. Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression. *IEEE Transactions on Reliability*, 62, 12 2013.

[119] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 130:377–388, 2017.

[120] C. J. Lu and W. O. Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.

[121] P. Luong and W. Wang. Smart sensor-based synergistic analysis for rotor bar fault detection of induction motors. *IEEE/ASME Transactions on Mechatronics*, 25(2):1067–1075, 2020.

[122] Y. Ma, Y. Lv, R. Yuan, and G. Song. Matching synchroextracting transform for mechanical fault diagnosis under variable-speed conditions. *IEEE Transactions on Instrumentation and Measurement*, pages 1–1, 2021.

[123] K. Mahesh Kumar and A. Rama Mohan Reddy. An efficient k-means clustering filtering algorithm using density based initial cluster centers. *Inform. Sci.*, 418-419:286–301, 2017.

[124] W. Mao, S. Tian, J. Fan, X. Liang, and A. Safian. Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation. *J. of Manufacturing Syst.*, 55(April 2019):179–198, 2020.

[125] N. Martin and C. Mailhes. Automatic data-driven spectral analysis based on a multi-estimator approach. *Signal Process.*, 146, 01 2018.

[126] K. McNaught and A. Zagorecki. Using dynamic bayesian networks for prognostic modelling to inform maintenance decision making. pages 1155 – 1159, 01 2010.

[127] K. Medjaher, D. Tobon Mejia, and N. Zerhouni. Remaining useful life estimation of critical components with application to bearings. *IEEE Transactions on Reliability - TR*, 61:292–302, 06 2012.

[128] W. Meeker, Y. Hong, and L. Escobar. Degradation models and analyses. 08 2011.

[129] G. Michau and O. Fink. Unsupervised fault detection in varying operating conditions. In *2019 IEEE Int. Conf. Prog. Health Manage.*, 06 2019.

[130] G. Michau, Y. Hu, T. Palmé, and O. Fink. Feature learning for fault detection in high-dimensional condition monitoring signals. volume 234, pages 104–115, 2020.

[131] J. Moubray. *Reliability-centered maintenance*. Industrial Press Inc., 2001.

[132] N. Nasios and A. G. Bors. Variational learning for gaussian mixture models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(4):849–862, 2006.

[133] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier. Pronostia: An experimental platform for bearings accelerated degradation tests. pages 1–8, 06 2012.

[134] W. Nelson. Accelerated Degradation. In *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, pages 521–548. 2008.

[135] F. Nelwamondo, T. Marwala, and U. Mahola. Early classifications of bearing faults using hidden markov models, gaussian mixture models, mel-frequency cepstral coefficients and fractals. *International Journal of Innovative Computing, Information and Control*, 2, 01 2006.

[136] H. Niaoqing, C. Min, and W. Xisen. The application of stochastic resonance theory for early detecting rub-impact fault of rotor system. *Mech. Syst. Signal Process.*, 17(4):883–895, 2003.

[137] T. Oberlin, S. Meignen, and V. Perrier. The fourier-based synchrosqueezing transform. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 315–319, 2014.

[138] A. P. Ompusunggu. Automated cepstral editing procedure (ACEP) as a signal preprocessing in vibration-based bearing fault diagnostics. In *Int. Conf. Surveillance 8*, Roanne Institute of Technology, France, 2015.

[139] S. Osman and W. Wang. A leakage-free resonance sparse decomposition technique for bearing fault detection in gearboxes. *Measurement Science and Technology*, 29:035004, 2018.

[140] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[141] B. Pang, G. Tang, C. Zhou, and T. Tian. Rotor Fault Diagnosis Based on Characteristic Frequency Band Energy Entropy and Support Vector Machine. *Entropy*, 20(12), 2018.

[142] C. Park, D. Moon, N. Do, and S. M. Bae. A predictive maintenance approach based on real-time internal parameter monitoring. *International Journal of Advanced Manufacturing Technology*, 85(1-4):623–632, 2016.

[143] P. Potočnik and E. Govekar. Semi-supervised vibration-based classification and condition monitoring of compressors. *Mechanical Systems and Signal Processing*, 93:51–65, 2017.

[144] G. Prakash, S. Narasimhan, and M. D. Pandey. A probabilistic approach to remaining useful life prediction of rolling element bearings. *Structural Health Monitoring*, 18(2):466–485, 2019.

[145] P. Primus. Reframing unsupervised machine condition monitoring as a supervised classification task with outlier-exposed classifiers technical report. 2020.

[146] W. Qian, S. Li, J. Wang, Y. Xin, and H. Ma. A new deep transfer learning network for fault diagnosis of rotating machine under variable working conditions. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pages 1010–1016, 2018.

[147] Y. Qu, D. He, J. Yoon, B. Van Hecke, E. Bechhoefer, J. Zhu, B. V. Hecke, E. Bechhoefer, B. Van Hecke, E. Bechhoefer, and J. Zhu. Gearbox tooth cut fault diagnostics using acoustic emission and vibration sensors–a comparative study. *Sensors (Basel, Switzerland)*, 14(1):1372–1393, 2014.

[148] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca. Condition-Based Maintenance—An Extensive Literature Review. *Machines*, 8(2), 2020.

[149] S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *SIGMOD Rec.*, 29(2):427–438, May 2000.

[150] R. Randall and N. Sawalhi. A new method for separating discrete components from a signal. *Sound and Vibration*, 45(5):6–9, 2011.

[151] R. Randall and W. Smith. New Cepstral Methods for the Diagnosis of Gear and Bearing Faults Under Variable Speed Conditions. *International Congress on Sound Vibration*, (July):1–8, 2016.

[152] R. B. Randall. *Vibration-Based Condition Monitoring*. Wiley, 2010.

[153] R. B. Randall and J. Antoni. Rolling element bearing diagnostics-A tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011.

[154] G. Ratsch, S. Mika, B. Scholkopf, and K.-R. Muller. Constructing boosting algorithms from svms: an application to one-class classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(9):1184–1199, 2002.

[155] D. A. Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.

[156] M. Z. Rodriguez, C. H. Comin, D. Casanova, O. M. Bruno, D. R. Amancio, L. d. F. Costa, and F. A. Rodrigues. Clustering algorithms: A comparative approach. *PLOS ONE*, 14(1):1–34, 2019.

[157] L. Ruff, R. A. Vandermeulen, B. J. Franks, K.-R. Müller, and M. Kloft. Rethinking assumptions in deep anomaly detection, 2021.

[158] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[159] H. Rupal Singh, S. R. Mohanty, N. Kishor, and K. Ankit Thakur. Real-time implementation of signal processing techniques for disturbances detection. *IEEE Transactions on Industrial Electronics*, 66(5):3550–3560, 2019.

[160] S. J. Russell , Norvig, Peter,,. *Artificial intelligence : a modern approach*. Pearson, Boston, 2020.

[161] A. Sadhu, G. Prakash, and S. Narasimhan. A hybrid hidden markov model towards fault detection of rotating components. *Journal of Vibration and Control*, 23(19):3175–3195, 2017.

[162] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel. Metrics for offline evaluation of prognostic performance. *Int. J. of Prognostics Health Manag.*, 1(1):1–20, 2010.

[163] A. Saxena and K. Goebel. Turbofan engine degradation simulation data set. NASA Ames Prognostics Data Repository, 2008.

[164] S. Schmidt, P. S. Heyns, and J. P. de Villiers. A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques. *Mechanical Systems and Signal Processing*, 100:152–166, 2018.

[165] J. Schoukens, Y. Rolain, G. Simon, and R. Pintelon. Fully automated spectral analysis of periodic signals. *IEEE Trans. Instrum. Meas.*, 52(4):1021–1024, 2003.

[166] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating support of a high-dimensional distribution. *Neural Comput.*, 13:1443–1471, 07 2001.

[167] B. Settles. Active learning literature survey. 2009.

[168] S. Shao, S. McAleer, R. Yan, and P. Baldi. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 15(4):2446–2455, 2019.

[169] H. Shi, L. Guo, S. Tan, and X. Bai. Rolling bearing initial fault detection using long short-term memory recurrent network. *IEEE Access*, 7:171559–171569, 2019.

[170] S. W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, USA, 1997.

[171] F. Sun, X.-Y. Li, T. Jiang, and X. Zhang. A bayesian least-squares support vector machine method for predicting the remaining useful life of a microwave component. *Advances in Mechanical Engineering*, 9:1–9, 01 2017.

[172] E. Sutrisno, H. Oh, A. S. S. Vasan, and M. Pecht. Estimation of remaining useful life of ball bearings using data driven methodologies. In *2012 IEEE Conf. Prog. Health Manage.*, pages 1–7, 2012.

[173] F. Swana and W. Doorsamy. An unsupervised learning approach to condition assessment on a wound-rotor induction generator. *Energies*, 14:602, 01 2021.

[174] X. Tang, L. Zhuang, J. Cai, and C. Li. Multi-fault classification based on support vector machine trained by chaos particle swarm optimization. *Knowledge-Based Systems*, 23(5):486–490, 2010.

[175] D. Tax and R. Duin. Support vector data description. *Mach. Learn.*, 54:45–66, 01 2004.

[176] I. Thoidis, M. Giouvanakis, and G. Papanikolaou. Semi-supervised machine condition monitoring by learning deep discriminative audio features. *Electronics*, 10(20):2471, Oct 2021.

[177] Z. Tian. An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring. *Journal of Intelligent Manufacturing*, 23(2):227–237, 2012.

[178] Z. Tian and H. Liao. Condition based maintenance optimization for multi-component systems using proportional hazards model. *Reliability Engineering System Safety*, 96(5):581–589, 2011.

[179] M. Tipping. Sparse bayesian learning and relevance vector machine. *J. Mach. Learn. Res.*, 1:211–244, 01 2001.

[180] V. Tran, H. Pham, B.-S. Yang, and T. Nguyen. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and

support vector machine. *Mechanical Systems and Signal Processing*, 32:320–330, 10 2012.

[181] A. H. C. Tsang. Condition-based maintenance: tools and decision making. *Journal of Quality in Maintenance Engineering*, 1:3–17, 1995.

[182] L. Uusitalo. Advantages and challenges of bayesian networks in environmental modeling. *Ecological Modelling*, 203:312–318, 05 2007.

[183] L. F. Villa, A. Reñones, J. R. Perán, and L. J. De Miguel. Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation. *Mechanical Systems and Signal Processing*, 25(6):2157–2168, 2011.

[184] M. Vishwakarma, R. Purohit, V. Harshlata, and P. Rajput. Vibration Analysis Condition Monitoring for Rotating Machines: A Review. *Materials Today: Proceedings*, 4(2):2659–2664, 2017.

[185] R. E. Walpole Myers, Raymond H., Myers, Sharon L.,, Ye, Keying.,. *Probability statistics for engineers scientists*. Prentice Hall, Boston, 2012.

[186] D. Wang, K. L. Tsui, and Q. Miao. Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. *IEEE Access*, 6:665–676, 2017.

[187] H. Wang, J. Xu, R. Yan, and R. X. Gao. A new intelligent bearing fault diagnosis method using sdp representation and se-cnn. *IEEE Trans. Instrum. Meas.*, 69(5):2377–2389, 2020.

[188] L. Wang and R. X. Gao. *Condition Monitoring and Control for Intelligent Manufacturing*. 2006.

[189] W. Wang. Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing*, 15(5):887–903, 2001.

[190] W. Wang. An enhanced diagnostic system for gear system monitoring. *Trans. Sys. Man Cyber. Part B*, 38(1):102–112, feb 2008.

[191] S. Watanabe. A widely applicable bayesian information criterion. *J. Mach. Learn. Res.*, 14(1):867–897, mar 2013.

[192] L. Wen, L. Gao, and X. Li. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1):136–144, 2019.

[193] Y. Wen, J. Wu, and Y. Yuan. Multiple-phase modeling of degradation signal for condition monitoring and remaining useful life prediction. *IEEE Trans. Reliab.*, 66(3):924–938, 2017.

[194] A. Widodo and B.-S. Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6):2560–2574, 2007.

[195] A. Widodo and B.-S. Yang. Application of relevance vector machine and survival probability to machine degradation assessment. *Expert Syst. Appl.*, 38:2592–2599, 03 2011.

[196] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics Intell. Laboratory Syst.*, 2(1):37–52, 1987.

[197] M. Xia, T. Li, L. Liu, L. Xu, and C. de Silva. Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder. *IET Science, Measurement Technology*, 11(6):687–695, Sept. 2017.

[198] J. Xie, L. Zhang, L. Duan, and J. Wang. On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis. In *2016 ieee international conference on prognostics and health management (icphm)*, pages 1–6. IEEE, 2016.

[199] D.-M. Yang, A. F. Stronach, P. Macconnell, and J. Penman. Third-Order Spectral Techniques for the DIagnosis of Motor Bearing Condition Using Artificial Neural Networks. *Mechanical Systems and Signal Processing*, 16(2):391–411, 2002.

[200] S. Yin, G. Wang, and H. R. Karimi. Data-driven design of robust fault detection system for wind turbines. *Mechatronics*, 24(4):298–306, 2014.

[201] J. Yu. Health condition monitoring of machines based on hidden markov model and contribution analysis. *IEEE Trans. Instrum. Meas.*, 61(8):2200–2211, 2012.

[202] J. Yu, X. Liu, and L. Ye. Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes. *IEEE Trans. Instrum. Meas.*, 70:1–15, 2021.

[203] S. H. Zargarbashi and J. Angeles. Identification of error sources in a five-axis machine tool using FFT analysis. *International Journal of Advanced Manufacturing Technology*, 76(5-8):1353–1363, 2014.

[204] C. Zhang, J. Jiang, W. Zhang, Y. Wang, S. M. Sharkh, and R. Xiong. A Novel Data-Driven Fast Capacity Estimation of Spent Electric Vehicle Lithium-ion Batteries. *Energies*, 7(12):8076–8094, 2014.

[205] Q. Zhang, C. Hua, and G. Xu. A mixture weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data. *Mechanical Systems and Signal Processing*, 43:103–112, 02 2014.

[206] Y. Zhang, C. Bingham, M. Gallimore, and J. Chen. Steady-state and transient operation discrimination by variational bayesian gaussian mixture models. In *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–5, 2013.

[207] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao. Deep learning and its applications to machine health monitoring. *Mech. Syst. and Signal Process.*, 115:213–237, 2019.

[208] X. Zhao, M. Li, J. Xu, and G. Song. Multi-class semi-supervised learning in machine condition monitoring. In *2009 International Conference on Information Engineering and Computer Science*, pages 1–4, 2009.

[209] Z. Zhao, T. Li, J. Wu, C. Sun, S. Wang, R. Yan, and X. Chen. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. *ISA Trans.*, 107:224–255, 2020.

[210] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta. Long short-term memory network for remaining useful life estimation. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 88–95, 2017.

[211] K. Zhu and T. Liu. On-line tool wear monitoring via hidden semi-markov model with dependent durations. *IEEE Transactions on Industrial Informatics*, PP:1–1, 07 2017.

[212] E. Zio and F. Di Maio. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear power plant. *Reliability Engineering System Safety*, 95:49–57, 01 2010.

APPENDICES

Appendix A

List of Publications

The following is a compilation of referred journal publications and conference papers directly resulting from the work contained in this thesis.

Publications in Refereed Journals:

1. **S. Fong** and S. Narasimhan. An unsupervised bayesian oc-svm approach for early degradation detection, thresholding, and fault prediction in machinery monitoring. *IEEE Transactions on Instrumentation and Measurement*, 71:1–11, 2022.

2. M. D. Kafle, **S. Fong**, and S. Narasimhan. Active acoustic leak detection and localization in a plastic pipe using time delay estimation. *Applied Acoustics*, 187:108482, 2022

3. **S. Fong**, J. Harmouche, S. Narasimhan, and J. Antoni. Mean shift clustering-based analysis of non-stationary vibration signals for machinery diagnostics. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4056–4066, 2020.

4. A. Ashasi-Sorkhabi, **S. Fong**, G. Prakash, and S. Narasimhan. A condition based-maintenance implementation for an automated people mover gearbox. *International Journal of Prognostics and Health Management*, 8(20):13, 2017.

Conference Proceedings - Full Paper:

1. **S. Fong**, S. Narasimhan, and M. Riseborough. A blind condition-based maintenance framework for real-time fault detection and degradation modeling of the link apm gearbox. In *Automated People and Automated Transit Systems 2020*, pages 43–54. 2020

2. **S. Fong**, A. Ashasi-Sorkhabi, G. Prakash, S. Narasimhan, and M. Riseborough. Automated condition-based maintenance of automated people movers. In *Proceedings of the 16th International Conference on Automated People Movers and Automated Transit Systems*, Tampa, FL, 2018. ASCE.

Appendix B

Application of S-MSC to Active Leak Detection in Water Distribution Networks

This appendix presents the extension of S-MSC to a new application outside of fault detection for rotating machinery: active leak detection and localization in fluid filled pipes. For leak detection and localization, S-MSC is applied as a spectral pre-processing tool to acoustic data obtained from hydrophone sensors installed inside fluid filled pipes. This appendix begins with a brief overview of the required theoretical background for active leak detection and localization, followed by a description of the experimental test bed and data set. Next, the S-MSC leak detection and localization procedure is presented, followed by results obtained using data collected from the experimental test bed.

B.0.1 Background

The theory and details presented here are solely intended to provide the reader with the necessary background to understand this particular use case for S-MSC. A detailed theoretical background and description of the work presented hereafter can be found in the co-authored publication [89].

In active leak detection, an active acoustic source located inside the pipe sends plane-wave excitations along the length of the pipe, which can be measured using hydrophone receivers located along the pipe. Physical discontinuities such as junctions, valves, as well as leaks, will cause a reflection in the portion of the incident the wave that comes

in contact with the impedance. This phenomenon manifests in the TF domain as a pair of same-frequency, similar amplitude peaks, separated by a time delay Δt, as illustrated in Fig. B.1. Using the theoretical or measured (i.e. experimentally determined) speed of sound in water, the measured time delays can be converted into an equivalent length of pipe which can be used to localize the position of the impedance or leak. For small impedances (i.e. leaks) the low energy of the corresponding peak pairs are easily buried within the noise of the signal. In addition, the transient nature of the excitation method coupled with other sources of time-varying phenomena (i.e. water usage) present in live, pressurized fluid filled pipes, the measured hydrophone signals will often contain many sources of non-stationarity which further complicates the task of peak pair extraction. Hence, in this application, S-MSC is used as a spectral pre-processing tool to extract low energy peak pair information buried in the signal noise.

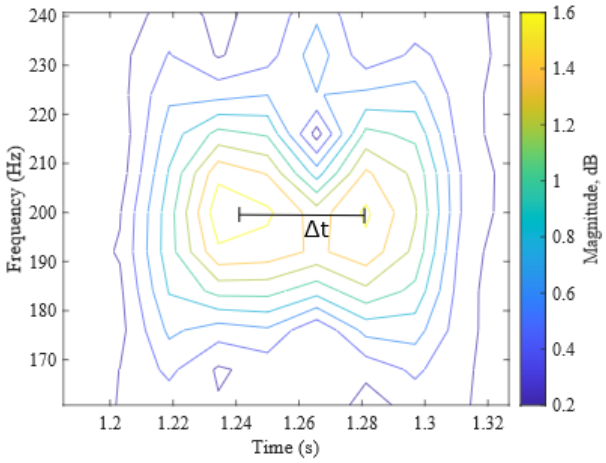

Figure B.1: Example TF representation of incident and reflected wave peak pair

B.0.2 Laboratory Test Bed

The data was obtained from experiments conducted on a straight 11.4m long pressurized PVC pipe located at the University of Waterloo. A diagram illustrating the configuration of the pipe, complete with labels for the hydrophone locations, leak location and other impendances is shown in Fig. B.2. Excitations produced by the acoustic source located on the end of the pipe are sampled at 25.6kHz using two Teledyne RESON TC4013 hydrophones (H1 and H2 in Fig. B.2 spaced 5.8m apart. The leak, located between the

hydrophones, is generated by opening a valve which simulates a leak with a flow rate in the range of 20L/min. To simulate realistic operating conditions, the pipe is connected to a pressurized water inlet (50-55 psi) connected to the City of Waterloo's main distribution system. The data analyzed hereafter pertains to experiments using a 0-100 Hz broadband excitation. Data collected separately for the non-leak and leak case.

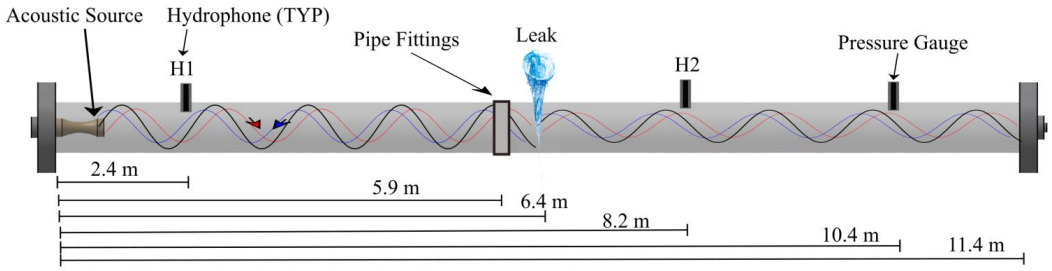

Figure B.2: Schematic of experimental set-up used for tests; also shown is the acoustic source, the hydrophone receiver locations (for H1-H2) and the leak location.

B.0.3 S-MSC Procedure for Leak Detection and Localization

S-MSC is used to perform leak detection and localization using hydrophone data obtained from a single sensor. The analysis pre-processing procedure is analogous to the approach presented in section 5.3. The key differences are in the post-processing steps follow S-MSC decomposition. Contrary to the approach used in the previous vibration analysis examples in which the noise spectrum of the S-MSC decomposition was discarded, this analysis will instead discard the harmonic spectra Z_c in order to obtain the time delay information buried inside the noise spectra Z_e. The post-processing procedure for leak detection and localization is as follows:

1. Using the non-leak data:
   (a) Use the S-MSC noise decomposition spectra Z_e to obtain a spectrogram representation of the "low-energy" frequency components in the original hydrophone signal.
   (b) Use peak finding to extract time delay information in the low-energy spectrogram

(c) Normalize and aggregate peak-pair information into a histogram representation with bin size equal to the STFT time resolution.

2. Calculate an upper statistical process control (SPC) limit for each histogram bin using the mean and standard deviation of the bin count obtained across all non-leak runs.

3. Using the leak data:

   (a) Repeat the process under step 1) for the leak hydrophone signals.

   (b) Detection: the bin frequencies of the leak histograms are compared against the non-leak control limits. A leak is detected when the upper control limit is exceeded at one or more of the histogram bins.

   (c) Localization: if a leak is detected in a bin, the time lag associated with that bin is converted into an equivalent distance from the hydrophone using the experimental wave speed ($\mu_c = 464m/s$)

B.0.4 Results

Fig. B.3a, b and c present an example of a the raw spectrogram and typical S-MSC high energy and low energy decomposition results, respectively, for H1 hydrophone data collected from the no-leak case. Fig. B.3a shows that the spectral content of the hydrophone signal is comprised of numerous low-energy regions which are interspersed between and obscured by higher energy regions. Fig. B.3c shows that the noise spectrum Z_e produced using S-MSC decomposition successfully unmasks the low energy signal components, allowing peak pair information to be extracted from the signal. An example histogram representation of the time lags extracted from the the S-MSC low energy spectrogram of a single pair of leak and non-leak tests is shown in Fig. B.4. The time lags associated with known impedances are labelled on Fig. B.4. Fig. B.4 shows that the largest discrepancy between the bin frequencies of the leak and non-leak histograms occurs at the leak location, for which the frequency is significantly higher in the leak case than in the non leak case. Similarly, the frequency of the bins for the leak case at the endcap, H2 and the pressure gauge decreases proportionally.

Detection and Localization

Normalizing and aggregating the time delay information across all non-leak runs results in the histogram and control limits shown in Fig. B.5. The superimposed leak histogram

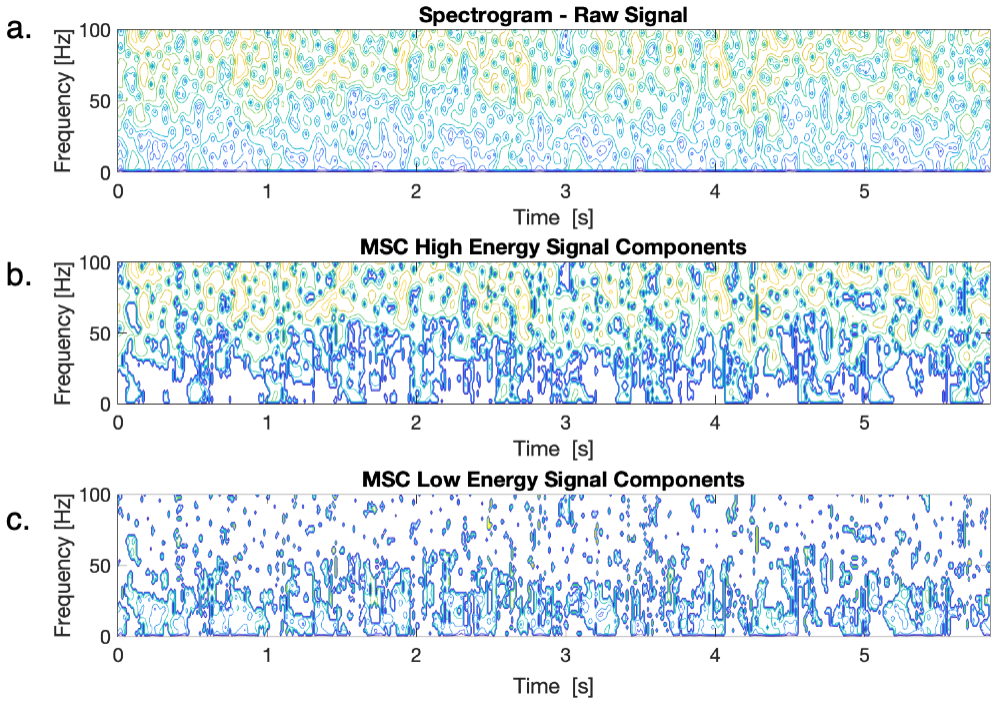

Figure B.3: Sample spectrograms for no-leak single pipe hydrophone: (a) raw signal, (b) MSC high energy components, (c) MSC low energy components

information is shown to exceed the non-leak SPC limit at only one bin: 0.0175s. Using the mean experimental speed of sound (464 m/s) to convert the time lags associated with the histogram bin center and edges corresponds to a leak location centered at 4.06 m measured from H1, bounded by [3.77m 4.35m]. Taking center distance of 4.06 m produces an estimate within 2% error of the actual leak location. The maximum error associated with the bin edges remains within 9% of the actual leak location. Alternatively, a probabilistic estimate of the leak location can be obtained by using the distribution of the experimental speed of sound instead - the corresponding result using this approach is shown in Fig. B.6.

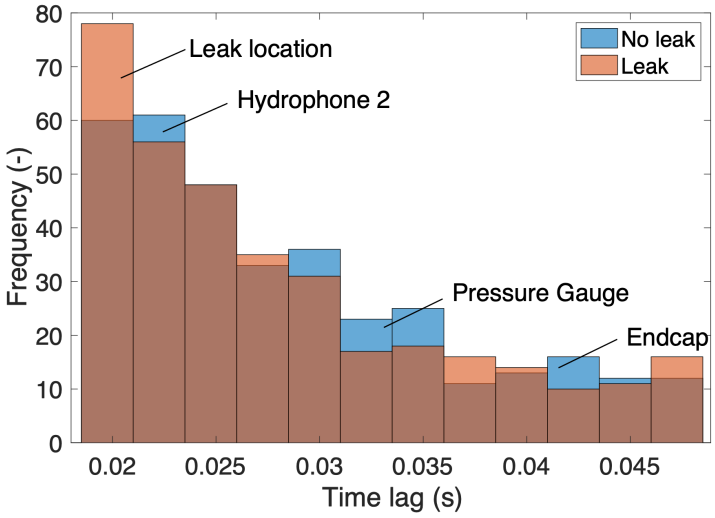

Figure B.4: Aggregated histogram of pairs extracted from MSC low-energy spectrogram with known impedances labelled - Hydrophone H1 (n = 350)

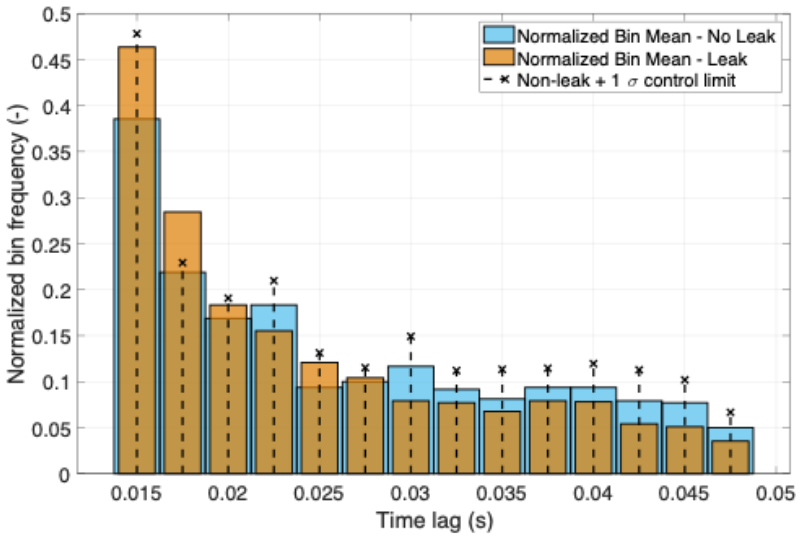

Figure B.5: SPC control chart based on normalized leak vs. no leak histograms - 100 hz bandwidth excitation - Hydrophone H1

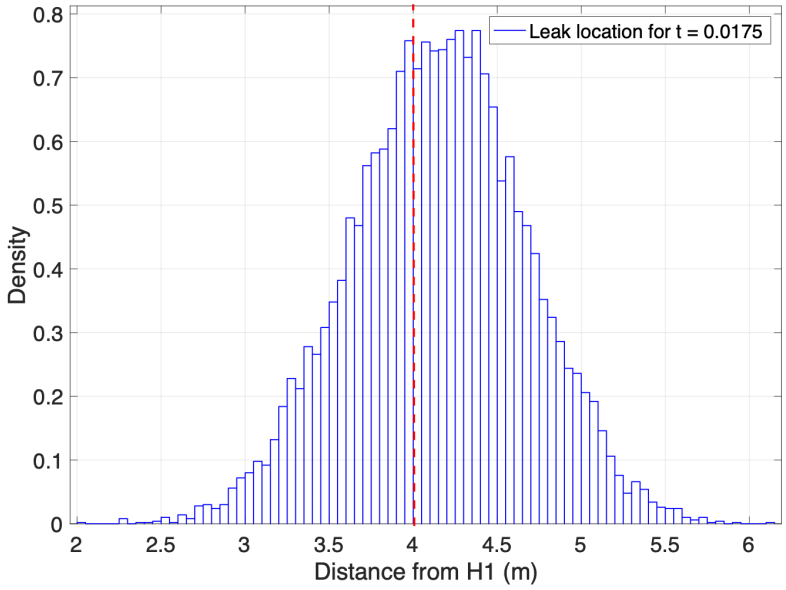

Figure B.6: Probabilistic leak localization result using SPC control chart histogram bin (t = 0.0175s) and statistical distribution of μ_c - true leak location labelled in red

Appendix C

Labelled Study on APM Operating States

In the preliminary phase of the LINK APM field pilot, a supervised study was conducted to determine a physical interpretation for the clusters found in the feature data. Vibration data collected from the APM gearbox was labelled with respect to the position of the train along its route. The following is a brief summary of the study results pertaining to accelerometer A1. The feature pair of hyper-kurtosis and crest factor used for the study were heuristically determined at the time.

Fig. C.1 (left) illustrates the position-labelled feature pair data, while Fig. C.1 (right) illustrates the corresponding result obtained from the minimum-BIC GMM model. Comparison of the two plots in Fig. C.1 shows that GMM clustering for this feature pair is consistent with the position/direction of the train. Further evaluation shows that clusters 1 and 2 correspond to different directions of travel for the train, while cluster 3 corresponds to scatter, containing points from all track segments.

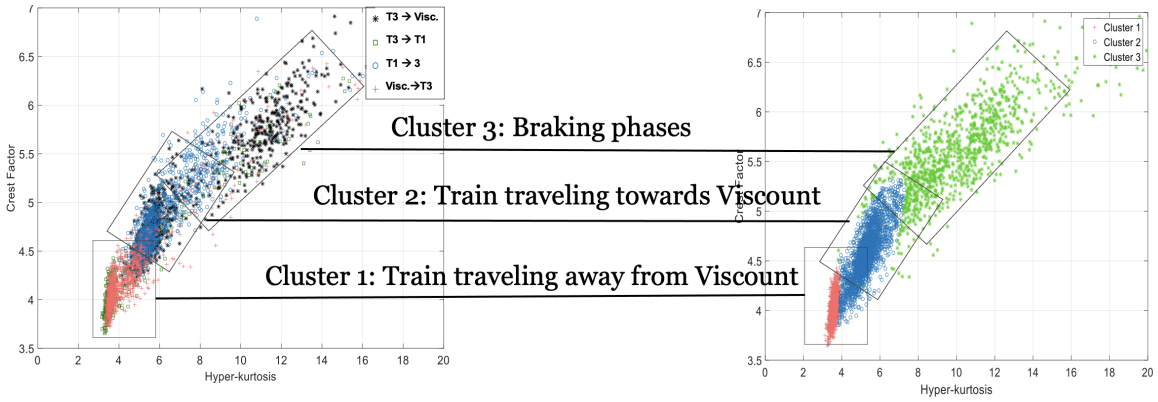

Figure C.1: GMM APM Comparison Example