Automated Fault Diagnosis in Rotating Machinery

by

Shilpa Reddy Pantula

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2014

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

Rotating machinery are an important part of industrial equipment. Their components are subjected to harsh operating environments, and hence experience significant wear and tear. It is necessary that they function efficiently all the time in order to avoid significant monetary losses and down-time. Monitoring the health of such machinery components has become an essential part in many industries to ensure their continuous operation and avoiding loss in productivity. Traditionally, signal processing methods have been employed to analyze the vibration signals emitted from rotating machines. With time, the complexity of machinery components has increased, which makes the process of condition monitoring complex and time consuming, and consequently costly. Hence, a paradigm shift in condition monitoring methods towards data-driven approaches has recently taken place towards reducing complexity in estimation, where the monitoring of machinery is focused on purely data-driven methods.

In this thesis, a novel data-driven framework to condition monitoring of gearbox is studied and illustrated using simulated and experimental vibration signals. This involves analyzing the signal, deriving feature sets and using machine learning algorithms to discern the condition of machinery. The algorithm is implemented on data from a drivetrain dynamics simulator (DDS), equipment designed by Spectraquest Inc. for academic and industrial research purposes. Datasets from pristine state and faulty gearboxes are collected and the algorithms are tested against this data. This framework has been developed to facilitate automated monitoring of machinery in industries, thus reducing the need for manual supervision and interpretation.

Acknowledgements

I take this opportunity to thank my supervisor Professor Sriram Narasimhan for giving me the opportunity and guidance to work on this thesis. I would like to express my gratitude to Dr. Budhaditya Hazra for mentoring me during the learning process and for providing valuable insights into the domain which made the process easier for me.

I thank Professor Scott Walbridge and Professor James Craig for taking their time to read my thesis and providing feedback. I thank Richard Morrison for helping us setup the DDS.

This work being part of a project in collaboration with Toronto Pearson International Airport, operated by Greater Toronto Airport Authority (GTAA), I would like to acknowledge the GTAA team for providing us valuable industrial perspective.

I would like to thank my group mates - Pampa Dey, Dr. Ayan Sadhu, Guru Prakash, Ann Sychterz and Kevin Goorts; and my office mate Atena Pirayehgar for their support and cooperation.

I thank my friends in Waterloo for all the good times and for everything that has been possible.

Finally, I thank my mom, dad, brother and my friend Tanuja Kambham for their support and motivation in completing my Masters. It would not have been possible without them.

Dedication

To my mom, dad and brother.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Rotating machinery constitute an important mechanical component of industrial infrastructure. Major rotating machinery applications include aircraft engines, automotive equipment, fans and blowers, turbines, industrial compressors, expanders and turbochargers, pumps and conveyor systems. The key common component of all the aforementioned machinery is the gearbox. Due to its continual nature of operation, an efficient and fault-free performance is a major requirement. Faults, especially if they are un-anticipated, can be costly and can cause significant financial losses. Furthermore, due to relatively harsh operating conditions, rotating machinery components are prone to early damage, leading to reduced service life of the operating unit or shutdown in severe conditions. It is thus imperative that the condition of machinery —in particular the gearbox —is monitored regularly. This research undertakes the problem of condition monitoring of gearboxes from a practical implementation point-of-view, using a combination of signal processing, condition

indicators, and machine learning algorithms.

While the gearbox is in operation, the assemblage of rotating parts (gears, shafts, bearings, etc.) generate vibration signals in various frequency bands, mostly lying in the human audible range (0 - 20 kHz). Traditional methods of diagnosis involve auditory supervision to detect familiar sounds from the machinery and inferring faults based on a recognizable acoustic pattern, leading to alerts pertaining to repair and replacement. Such a system based on human perception is fraught with uncertainties and risk of higher incidence of false positives in case of heavy and complex machinery. This underscores the requirement of developing sophisticated tools for condition monitoring of rotatory machinery. The vibration signals are commonly analyzed through sophisticated signal processing algorithms to detect faults, some of which are described by [58].

Gearbox signals are primarily composed of rotational harmonics and meshing harmonics, including their overtones, alongside their fundamental frequency [21, 61]. In its pristine state, gear signatures resemble a sum of sinusoids with frequencies that are integer multiples of the fundamental gear meshing frequency. As the gears deteriorate, sidebands start to appear in some of the harmonics and thus the signal becomes more of a sum of sinusoids with amplitude and frequency modulation (AM-FM). Vibration signatures associated with rolling element bearings consist a periodic series of ringing pulses resulting from elements rolling over a sharp edge, crack, or chip [53, 61]. Thus, the energy is spread across a wide band of frequencies that could be easily masked in the presence of signals generated by imbalance, misalignment, gear meshing, etc. To cater to the complexities introduced by the combined presence of gear, bearing and shaft vibration components, it is necessary to use sophisticated signal pre-processing tools to clean the data, or to extract relevant components that are easy to process through the data driven methods. Figure 1.1 shows a scenario when two gears mesh and the corresponding signature in terms of frequencies.

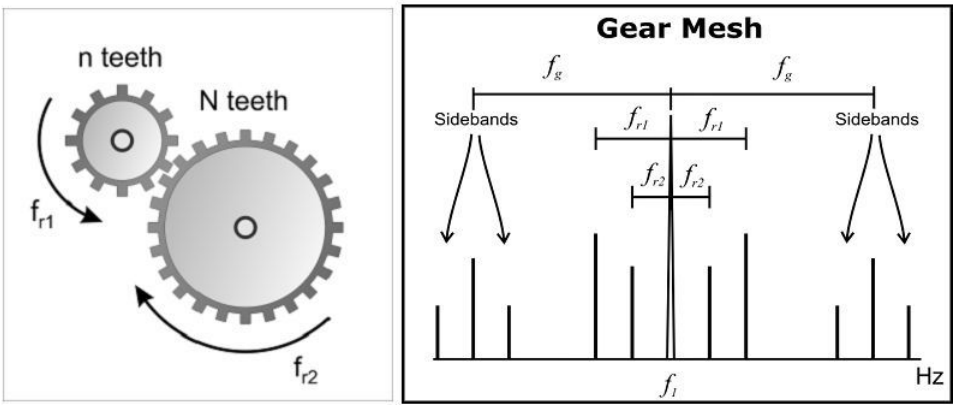


Figure 1.1: Gear Vibration Signature [6]

Figure 1.2 shows the vibration signature from a bearing that had a fault in the outer race.

It is important that the methodology developed for fault diagnosis balances simplicity of implementation with the complexity of gearbox vibration signals. Most commonly used data driven approaches rely on information present in the vibration data and not on the configuration of the gearbox, as presented by [68]. In these methods, condition indicators (CI) calculated from vibration signals are used as feature set [62]. Data driven methods derive similarity between fault cases and analyze patterns in the vibration data. In addition, they can also be used to reduce the dimension of the data thereby enabling better comprehension and representation [11]. When used with novelty detection [67, 70], they can detect faults in an expedient fashion. In the present study, the data-driven algorithms are augmented with some basic system level information from the gearbox (e.g., meshing harmonics) and formulated to simplify the overall condition monitoring process.

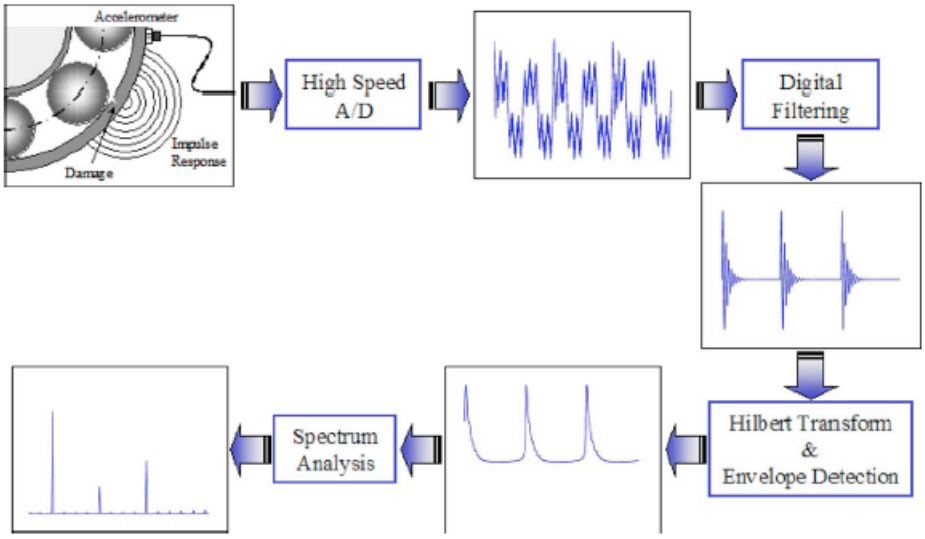


Figure 1.2: Bearing Vibration Signature [43]

1.2 Objectives

The main goals of the thesis are as follows:

1. Develop a systematic automated methodology to diagnose faults in rotating machinery, specifically gears, using vibration data.

2. Build, test, and demonstrate the developed approach using experimental test data acquired from a drive-train diagnostics simulator.

1.3 Overview of Approach

In simple terms, the central objective in this thesis is to detect faults in gear motors from vibration signatures. This diagnosis is undertaken for different fault conditions, while retaining the same operating speed for all the cases. A typical gearbox contains both gears and bearings and both these components are important for fault detection in a gearbox. However, the focus of this thesis is on gear damage. To isolate the gear mesh frequencies (GMFs), a signal processing technique called empirical wavelet decomposition (EWD) is used to filter out the unnecessary frequency components from the signals [26]. Condition indicators are then applied to the processed data, which contain information about the signal variation between fault cases. Since the number of condition indicators is large and not all the features contain useful information, the dimension of the condition indicators (also called feature space) is reduced using principal component analysis (PCA) and visualized using a scatter diagram. For automated inference, the CIs are processed using a novelty detection tool.

For novelty detection, Mahalanobis distance (MD) is used as a similarity measure between the data (specifically, CIs). As the data is being acquired, a process monitoring tool called statistical process control (SPC) analyzes the incoming data and alerts when it detects a change. The logarithm of MD is used as process control variable. Simultaneously, online clustering is performed on the data using a Gaussian mixture model (GMM).

This fault detection framework is implemented on the data obtained from a bench-scale test set-up called a drive-train diagnostics simulator (DDS), designed by Spectraquest Inc. The DDS contains a two-stage parallel shaft gearbox, roller bearings, a magnetic brake and is driven by a variable frequency drive induction motor. This set-up simulates a suite of fault cases of a gearbox using different types of test samples of gears and roller bearings

(explained in Chapter 4). In the data-acquisition stage, this algorithm is implemented on an NI CompactRIO set-up in the realtime mode to analyze the results.

1.4 Organization

Chapter 2 introduces the basic concepts on which modern day machine diagnostics procedures are based. This Chapter also reviews literature present in the area of diagnostics of rotating machinery. It also explains the importance of data driven methods for diagnostics and prognostics.

Chapter 3 describes the methodology developed in this thesis for machine diagnostics. It describes the application and limitations of the method.

Chapter 4 illustrates the operation of the DDS and describes implementation of the algorithm on data from the DDS. As well, the CompactRIO system implementation in realtime to process the data is described in the Chapter.

Chapter 5 concludes the thesis by summarizing major findings and proposes ideas for future research.

Chapter 2

Background

Some rotating machines are operated continuously in harsh environments and are prone to rapid deterioration. It is desirable that continuous monitoring systems be implemented to monitor their condition in real time. This Chapter reviews existing methods for condition monitoring of machinery with a focus on automating condition monitoring of gearboxes [51]. A gearbox uses gears and gear trains to provide speed and torque conversions from a rotating power source to another device [60]. It also consists of bearings that are critical in smoothening the movement between gear shafts and fixed ends. In Section 2.1, a review of traditional methods for condition monitoring of gearboxes is provided and their drawbacks are described. Methods developed with the objective of improving the condition monitoring of gearbox are reviewed in Section 2.2. Section 2.3 describes the limitations posed by existing condition monitoring techniques, some of these are addressed in this thesis as explained in further Chapters.

2.1 Literature Review

2.1.1 Condition Monitoring of Gearboxes

Maintenance of machines is as important as their installation and operation for efficient performance of the industrial equipment. Three commonly used strategies for maintenance of machinery include run-to-break, time-based preventive maintenance, and condition-based maintenance (CBM)[7, 65]. Run-to-break is a traditional method where machines are run to failure. This results in the longest operating time of machines and could potentially lead to the maximum damage, induced at failure which in turn could lead to increased down-time and cost to repair and production costs. However, this strategy is best applicable in industries where the machines are small, wherein the risk of failure is minimal and cost of replacement/repair is less. Time-based preventive maintenance is a policy wherein machines are regularly monitored such that the time between inspection is less than the time between failures. This method is suitable where the time of failure is predictable and can lead to prevention of failures. But this approach ignores the possibility that not all components have a predictable failure rate, which could lead to fatal damage. So, it is not suitable for components whose lifetime cannot be predicted with confidence. CBM is based on predictive maintenance, in that the time to failure of machine is predicted based on the current behaviour of the machine. This method of condition monitoring is efficient in reducing maintenance costs and at the same time increases the operating life of machines.

This thesis primarily deals with methods for CBM of machinery. Research in condition monitoring has developed due to relatively large industrial demand, and hence a large body of literature is available [32, 52]. Condition monitoring of machinery involves planning

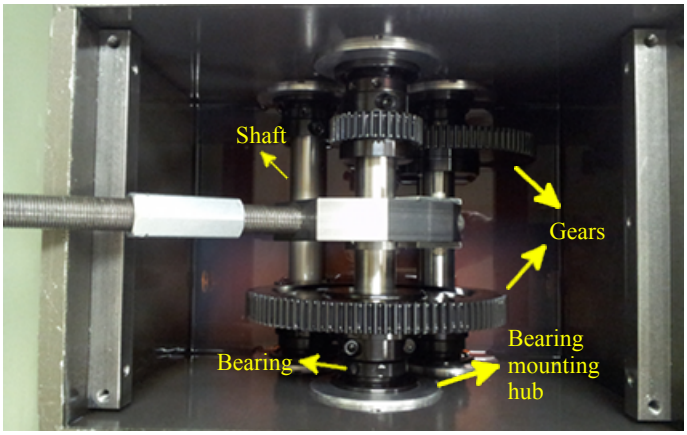


Figure 2.1: Components in Gearbox - Gears and Bearings

a maintenance schedule based on the current condition of the machinery. In order to achieve this, the maintenance strategy should analyze the condition of machinery while it is operating. Two important ways of obtaining such information are vibration analysis and lubricant analysis; others include performance analysis and thermography [51]. Vibration analysis is based on studying vibrations generated by the machine during its operation. A fault developing in the machine is reflected in its vibration signatures; thus, information from analyzing the vibrations will describe the intrinsic fault condition. Lubricant analysis entails processing the lubricant that carries information in the form of wear particles, chemical contaminants, etc. The condition of some machines can be discerned from the lubricant. In the present study, the main focus is limited to vibration based CBM and hence the other aforementioned methods will not be discussed.

A gearbox typically contains gears and bearings aligned on a shaft that rotates, the rotation is transmitted through gears and bearings (see Figure 2.1). A machine in any condition generates vibrations during operation. When the shaft inside a gearbox rotates, there are frictional and rotational forces generated. The vibration created by these forces

gets transferred through the bearings to the gearbox housing. Events happening inside the gearbox like rotating shafts, meshing gear teeth, rotating electric fields, etc, are periodic and the vibrations are linked to these events. The periodicity of these events' occurrence describes the source of vibrations, thus the vibration analysis is mostly based on frequency analysis [51]. There is another category of vibrations that are generated due to fluid flow. The third category of vibrations are generated from torsional vibrations due to angular velocity fluctuations of shafts and other components. All these vibration signatures carry information about the condition of the machinery.

This research deals with vibrations signatures acquired using accelerometers (i.e. acceleration signals) that are mounted at certain key locations on the machine. Vibration analysis is widely used in machine condition monitoring because of the advantages it offers over other methods of condition monitoring. It can detect faults immediately because it reacts immediately to changes and after processing the signals, even slight indications of fault can be tracked. In comparison to oil analysis, vibration analysis performs efficiently because a minor fault in machine will not cause changes in its chemical composition, but will increase the intensity of vibrations which are easy to detect [51]. This research is carried out using vibration analysis because of the advantages it offers for condition monitoring of gearbox and for automating the process.

To measure vibrations from machines, a set of transducers are placed and standard procedures are followed [20]. Transducers are used to measure vibrations from machines in mechanical form and convert them into electric signals [51]. The most commonly available types of transducers are displacement, velocity and acceleration transducers [51], force transducers and torsional vibration transducers. There are many types of vibration transducers depending on the type of vibrations they measure - proximity probes, velocity transducers, acceleration transducers, dual vibration probes and laser vibrometers. Proximity

probes measure relative displacement between the probe tip and the surface on which it is mounted [51]. Velocity transducers measure signals proportional to velocity. Acceleration transducers, also called accelerometers, measure signals proportional to acceleration. The internally amplified piezoelectric type of accelerometers are commonly used. The charge output generated by the sensor is proportional to the force and therefore acceleration. In this case an amplifier is needed to convert the charge output into a voltage output and the amplifier is powered by the data collector [51]. The advantage of accelerometers is that they measure vibrations in a wide frequency and amplitude range and are very stable because they maintain calibration for a long time. Shaft vibration is usually measured by proximity probes like encoders and tachometers.

In this research, Dytran accelerometers are used. Mounting accelerometers properly is important to obtain good vibrations from the machine. Because they are sensitive to mounting techniques and surface conditions, their installation has to be identified accurately before testing. Different practices of mounting accelerometers are shown in Figure 2.2 [9].

2.1.2 Signal Processing Methods

Signal processing methods are a key component of the CBM of gearboxes. At the heart of the problem lies our ability to resolve changes due to the health of the system from changes in normal operating conditions. This presents significant challenges as vibration signals are mostly polluted with noise, and the system dynamics are only approximately known. Hence, correctly diagnosing small changes is challenging. Added to this, gearbox signals are inherently complex due to contribution from a large number of moving parts. Thus, a wide range of signal processing techniques have been developed in the last two decades

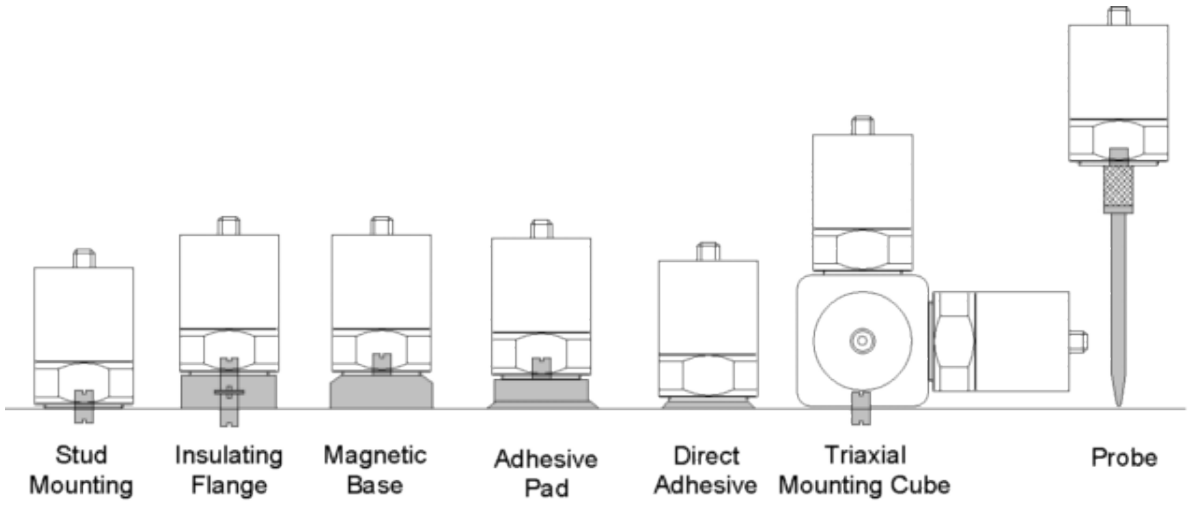


Figure 2.2: Mounting Accelerometers

[32], a complete account of which is not possible within the scope of this work and only a brief review is attempted here.

In the classical approach for CBM, vibration signals are considered in their raw form and signal processing techniques are directly applied on them with the application of signal enhancement techniques [32]. These techniques utilize methods for power spectrum estimation, fast Fourier transform (FFT), cepstrum analysis, and envelope spectrum analysis, etc., and have been found to be effective in gear fault detection. However, these methods are based on the assumption of stationarity and linearity of the vibration signal and hence are limited in their application. Gear fault signatures are time-localized transient events and hence non-stationary by nature. At an early stage, faults manifest themselves as impulsive events and for an early diagnosis it is necessary to utilize methods aimed at tracking frequency content [73].

Dealing with non-stationary and nonlinear signals requires the use of time-frequency

analysis techniques such as the Short-Time Fourier Transform (STFT) [12], Wavelet Transform (WT) [54, 64] or Wigner-Ville distribution (WVD) [55]. Continuous Wavelet Transform (CWT) [12, 13] has been successfully used in non-stationary vibration signal processing and fault detection. Filter bank implementation of WT, namely the discrete wavelet transform (DWT) and wavelet packet transform, have also been successfully applied [13, 47]. The time frequency representations belonging to Cohen's class [1, 12] such as the Wigner-Ville distribution (WVD), Choi-Williams distribution [1] and their enhanced derivatives possess attractive features that makes them suitable candidates for gear fault diagnostics [38, 39]. However, the performance of WVD and CWD can be seriously impeded by the presence of cross terms, which is indicative of some spurious frequency components which can adversely affect the interpretation of the resulting T-F distribution.

In parametric approaches towards gear fault diagnostics, time series models have been applied to vibration signals analysis of rotating machinery, where the vibration signals are modelled using time-invariant coefficients [3, 18, 63]. Another powerful tool is the vector autoregressive model (VAR) which balances complexity and speed of computation. Since vibration signals are non-stationary, the coefficient matrices of VAR model are made to vary with time as well. Towards that end, the use of Kalman filtering, noise adaptive Kalman filtering, and extended Kalman filtering for modelling time varying vector ARMA models is noteworthy in gear fault diagnostics [73]. But all the parametric models suffer from one issue; the choice of model order. In multi-component gearbox signals corrupted with noise, the model order becomes quite large, and discerning a faulty from a healthy condition can be quite confounding.

In the family of modern methods, enhancement of the raw signal is undertaken to improve detection and to reduce false positives. The use of blind source separation (BSS) to separate useful components from rotating machine signals has witnessed widespread appli-

cations [10, 25, 72]. However, in practical problems only a handful of sensors (may be one or two) are available, and the performance of BSS in separating meaningful sources is questionable as the number of meshing and rotational harmonic components can easily surpass the number of sensors. Signal decomposition methods like empirical mode decomposition employed on single channel measurements have also found applications in gear-fault diagnosis [31, 49]. The main advantage of such techniques is the ease and simplicity of the approach. Empirical mode decomposition in particular is a powerful tool that is applicable to a majority of signal types encountered in practice. The main downsides of EMD include its ad-hocism, lack of mathematical structure, inability to separate closely spaced modes even with the use of linear filters, mode mixing and poor noise performance.

A new method called empirical wavelet decomposition (EWD) has recently been proposed [15], which effectively integrates the decomposing power of EMD and the richly endowed mathematical structure of wavelets, while making the filtering process more adaptive. This method can decompose noisy and non-stationary signals into components and provides an attractive alternative to EMD for gearbox signals [3]. Its main advantages include robustness in the presence of noise and no requirements of band pass filtering or intermittency criterion. In this study, empirical wavelet decomposition (EWD) [26] is used to isolate the region of interest in the proximity of gear mesh frequencies (GMF) by filtering out the unnecessary frequency components from the signals.

2.1.3 Drawbacks of Traditional Methods

1. The direct methods often do not yield accurate results, especially when the baseline data is unavailable or lacking. Furthermore, since they are applied on the signals directly, their performance is often compromised by the presence of noise and the

complexity of the signals themselves.

2. Applications of parametric methods often lead to prohibitively large values of model orders making the fault detection process resource consuming and error-prone. For example, estimating noise covariances using Kalman approaches is fraught with difficulty.

3. The performance of BSS in separating meaningful sources from complex gearbox signals can sometimes be unreliable, especially for the underdetermined case, where the number of sensor measurements available in practice is less than the dominant harmonics (number of meshing and shaft harmonics can be much greater than the 2 or 3 sensors typically used) present in the system.

4. The main concerns with the use of EMD are: lack of a proper mathematical structure or its empiricism, inability to separate closely spaced modes, mode mixing and poor performance in noisy conditions. Frequently, successful separation of sources using EMD requires the application of band pass filtering or intermittency criteria, which may not be practically feasible.

5. Most of these methods (with the exception of direct methods) require significant user intervention to study frequency content, and are generally not useful for automated diagnostics.

6. Signal processing methods only generate certain diagnostic patterns and they need to be processed by inference tools like pattern recognition, novelty detection, HMMs, to estimate the extent of faults or to classify different types of faults.

2.2 Recent Developments

Traditional methods based on signal processing and detection techniques are not adequate for addressing the problem of CBM. Visual representation of patterns generated using signals or comparison based on spectral or time frequency plots can lead to escalation of detection errors and increasing rate of false positives. Thus, in recent times there has been a shift towards combining several signal processing methods with the smart use of condition indicators, machine learning, and statistical process control not only to detect faults but also determine their extent. In this Section, recent developments on condition based maintenance are reviewed.

Variation in the gearbox condition could be because of a change in operating speed, loading condition, or development of faults. This variation in the condition of gearbox causes variation in the level of vibrations measured by accelerometers. This is critical for fault detection in a gearbox because these vibration signatures transmit information about the health of the gearbox prior to any fault. So, by comparing accelerations at different times, the condition of gearbox can be discerned. Processing these signals can provide important information about the gearbox (as described in Subsection 2.1.1). It is noticed that comparing the signals at different times alone cannot describe the fault conditions properly, so features have to be extracted from the vibration signals, which contain more information [22, Chapter 7], [36]. These features are used as condition indicators for monitoring machinery.

In the literature, three main categories of features have been noted and can be categorized based upon their domain of measurement: time domain, frequency domain, or time-frequency domain. Root mean square (RMS), kurtosis, skewness, normalized sixth moment, crest factor, standard deviation, peak factor, autoregressive (AR) parameters are

some examples of time domain metrics [22, 62, p. 173]. Peaks of FFT, power spectrum, frequency band energy are important frequency domain features, and wavelet coefficients is a common time-frequency feature [23]. These features may be useful indicators of system condition. Behaviour of a machine can be analyzed using such features in conjunction with machine learning algorithms [22].

Some features are highly sensitive to damage and they can indicate damage during very early stages, while some features may not be as sensitive. Sensitivity of features depends on both the prevailing environmental and operational conditions. While calculating features that are insensitive to external factors, there will be trade off between sensitivity and fault detection capabilities [14]. Bartelmus and Zimroz proposed new features for different type of diagnosis performed on the gearbox [4]. They use load susceptibility for discerning the condition of the gearbox and instantaneous input speed as an indicator of operating conditions [4]. Features are found to be useful in analyzing the vibration data for fault diagnosis without much information about the actual condition of the gearbox [56]. Features are also used for fault diagnosis using acoustic emissions because they indicate variation in condition of gearbox [59].

2.2.1 Reducing the Number of Features

Features obtained from vibration data can be lot more than that is required from information content point of view [5]. Moreover, with an increase in the number of features, the computational complexity of the algorithm for monitoring systems increases, often referred to as the curse of dimensionality [19, 22]. There are two ways of reducing the dimensionality in data - selecting subsets of features or extracting features. Selecting features is choosing those features expected to be most relevant to the problem from literature and

study. This method of dimensionality reduction does not always gives effective results because depending on the application, different features might contain more relevant information for processing with monitoring systems [19]. Thus, it is preferable to use the second method of reducing dimensions - feature extraction.

Feature extraction uses a linear transformation of the features to find a subspace representation of the features where discerning information becomes easier. Component analysis is finding a projection using least squares method, such that the components are best represented [19]. It seeks directions in which representation is efficient. An alternate to component analysis, discriminant analysis also uses least squares method for finding projection that best separates the data. It seeks directions in which the separation is efficient [19].

Timusk *et al.* perform feature selection between AR parameters, time domain and features calculated by resampling the time waveform of the vibration data [58]. A commonly used method of transforming features is the principal component analysis (PCA) [30, 41, 45, 68]. Subspace methods along with kernel PCA for gearbox fault detection are efficient for separating the data and for reducing dimensions [29]. To monitor a planetary gearbox in non-stationary conditions, PCA and canonical discriminant analysis (CDA) were used and it was observed that CDA performs better [74].

2.2.2 Machine Learning

Computing feature space provides an idea of the condition of the gearbox, but the next stage of fault identification namely, machine learning, provides the necessary evaluation criteria for decision making. If the fault conditions are known *a priori*, it is called supervised learning and if the fault conditions are unknown, it is unsupervised learning [22]. Both the

methods of learning have numerous applications in the fault diagnosis of machinery [69]. There are parametric and non-parametric approaches for machine learning. Parametric methods are based on statistical representation of the features, and by doing so, lower probability features are mapped to fault conditions because their occurrence is for shorter time in the data sets. Non-parametric methods include nearest neighbours type and neural network based methods [19].

Pattern recognition methods mainly focus on data representation where the similarity between data is denoted by the distance between them when they are plotted. Thus, closely spaced data denotes that they are from nearly identical fault condition of the gearbox. Distantly spaced data means that they are from different fault conditions. To denote the notion of closeness, a distance metric is vital. Euclidian distance and Mahalanobis distance [44] are commonly used metrics while others such as city block distance exist in literature. Mahalanobis distance is used in this research because it is independent of the scatter in the data, unlike the Euclidean distance [16, 69].

Using supervised learning on the features is a classification problem. Using the training data, the algorithm generates a model with the help of fault types present and compares incoming data with the model, thus assigning a fault type to the data. Generally, data is pre-processed using PCA and then classified into its fault cases with the help of labeled data [41]. Parameters such as minimum Euclidean distance and generalized Euclidean distance, or Bayes classifiers [19] are used to classify the data by calculating the probability that the data belongs to a particular class. Many studies are performed using non-parametric methods of classification because of their ability to handle the random nature of the data. Methods such as radial basis functions (RBF), neural networks, support vector machines (SVM) and multi layer perceptron (MLP) are widely used for non-parametric classification of gearbox data [8, 48, 50]. Nearest neighbour classification has also been used to classify

gearbox data [2].

2.2.3 Novelty Detection

For gearbox condition monitoring, unsupervised learning has witnessed widespread usage, primarily as a novelty detection tool. Novelty detection is identifying if all data points behave identically or if there is an abnormal behaviour. It is a one-class classifier, i.e., it classifies all possible data points into one class and the unfamiliar ones are left unclassified to be presumed as novel. The latter indicate that the gearbox is starting to behave abnormally or a fault is impending [22, Chapter 10]. This principle is the central idea used in the present work.

The principle of novelty detection is illustrated using an example. A dataset A of 100 points, whose mean M_A is given by $\begin{bmatrix} 1 & 1 \end{bmatrix}$ and covariance matrix C_A is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and dataset B containing 20 points with parameters $M_B = \begin{bmatrix} 4 & 4 \end{bmatrix}$ and $C_B = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ are considered, as shown in Figure 2.3. In this figure, the axes correspond to the two columns of simulated data. The blue circle points represent class A and the green square points represent class B. The red boundary represents the contour at 2 standard deviations (approximately 95^{th} percentile, when the data is modelled as Gaussian distribution). The figure shows that the points belonging to class B lie outside the boundary indicating that they don't belong to the class A.

To automate the process of novelty detection for condition monitoring, a quantity has be defined that allows for process monitoring and defining a process variable. In this work, logarithm of Mahalanobis distance (MD) from centroid of cluster corresponding to pristine

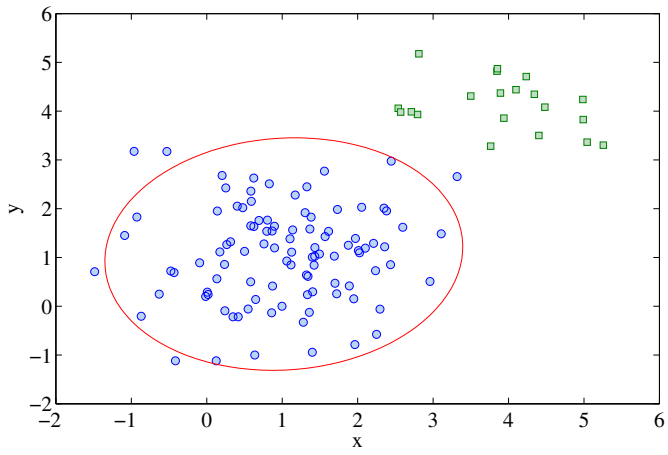


Figure 2.3: Distribution of Two Datasets

condition (see Appendix D) is used. The data points in class A belong to pristine condition of the machinery, so a threshold is set using the class A. $\log(MD)$ is plotted against the index of data points, and the datapoint from where the value of $\log(MD)$ is above the threshold, they are considered to be belonging to a novel class. In the Figure 2.4, the first 100 points belong to class A, so a threshold is set using the MD of first 100 points. As the points after 100 (which belong to class B) lie beyond the set threshold MD, the $\log(MD)$ corresponding to those 20 points lies above the threshold (represented by red dotted line). This principle of novelty detection as a one-class classifier is used to detect changes in the health condition of machinery and is elaborated in next Sections for automated fault diagnostics.

Timusk *et al.* [58] illustrate the use of novelty detection to detect abnormal behaviour in machinery. Unsupervised learning or clustering algorithms are used to define models for the pristine data. Classifiers like SVM, gaussian mixture models (GMM), self-organizing maps (SOM), neural networks and nearest neighbours are then used to cluster the pristine

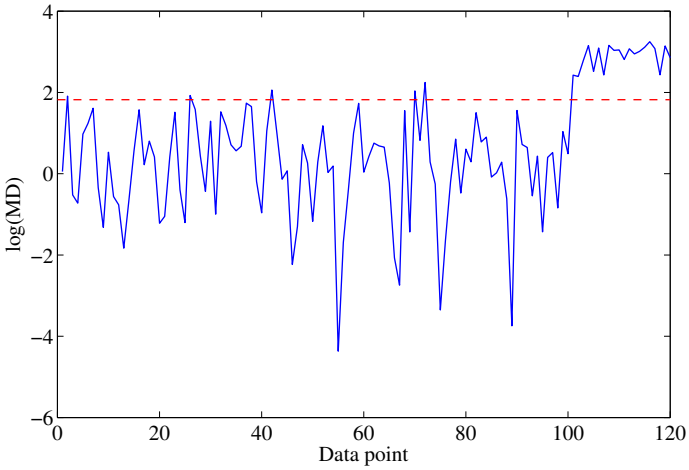


Figure 2.4: Comparing the Novelty Score against a Threshold

data and define thresholds using the parameters of the cluster containing the pristine data. Any incoming data that lies beyond the threshold is presumed as coming from a faulty machine indicating that the machine has developed faults.

In neural network based methods like SOM (see Appendix B), a reconstruction error is calculated based on the model created for the pristine data, and this error is used as a metric to set thresholds [28, 66]. Extreme value statistics are used in many applications to set thresholds for novelty detection [22]. An evolving novelty detection algorithm based on gaussian mixture fuzzy models has also been proposed for detecting incipient faults in machinery for automation [24].

2.2.4 Statistical Process Control

For industrial applications, it becomes necessary to automate the novelty detection process, reducing the need for manual inspection of data. Statistical process control (SPC)

provides a platform to achieve this automation [68]. SPC is implemented in two phases: first establish the process i.e., define process variables, and next define the monitoring control rules [46]. Using the process variables definition SPC charts such as Shewhart T^2 control charts, cumulative sum (CUSUM) control charts and exponentially weighted moving-average (EWMA) control charts give the output of the process [22]. When the process deviates from the control rule, an alert is issued.

This approach is very effective in that it enables early detection of faults when applied to gearbox condition monitoring. Traditionally in vibration analysis, the vibration signal itself is used as the process variable and the control limits are set based on the vibration signal for EWMA charts. However, it sometimes can be unstable and the vibration signal has to be transformed from its original state to be used as a process variable. The Hotelling or Shewhart T^2 chart uses Mahalanobis squared distance as the process variable. Mean μ and variance σ^2 are calculated using the process variable [22]. For example, AR(30) residual errors are used as process variables for X-bar chart in a study [22]. A multivariate SPC method based on independent component analysis has also been studied where the components are used as process variables [34].

For automated novelty detection approach in industry, a SPC framework was proposed by Filev *et al.* [23]. This method accounts for multiple operating conditions and the condition indicators are clustered using a modified Gaussian mixture model based on fuzzy logic. Novelty direction is applied on the Mahalanobis squared distance to detect abnormal behaviours in the machine. An EWMA chart is plotted using μ and σ^2 as variables, automatically detects a change, and reports it as a novel condition. A novel condition in this case could be a new operating condition of the machine or a fault condition. In order to distinguish between an operating condition and a fault, an algorithm has been proposed in literature [23] that distinguishes between new operating condition and an incipient fault.

This approach has myriad applications in industry because it not only enables automated condition monitoring, but also considers multiple operating conditions of the machinery [23].

2.3 Limitations in Existing Work

Existing methods of fault diagnosis are mostly based on signal processing and finding the frequency content in the signals that correspond to fault conditions. The spectral kurtosis and envelope analysis methods compare the frequency content of pristine and faulty machine signals. These methods are sensitive to noise. Analysis is performed offline after acquiring the data from gearboxes. This is associated with computational complexity and cost of labor. Besides, these systems are capable of diagnosing faults after they have occurred because the signals from faulty machinery is used to compare against the baseline data from machinery in pristine condition. An efficient fault diagnosis algorithm is required to diagnose faults at early stage of failure to avoid loss of time and money and should be installed to detect faults with least manual intervention.

2.4 Contributions of this Work

The algorithm and techniques presented in this research address some of these issues of fault diagnosis. The approach is less dependent on signal processing and more on data-driven approaches that require minimal manual intervention. These algorithms are based on unsupervised learning, making them easier to implement in cases where the baseline data is unavailable for training. Statistical process control approach is sensitive to changes in the data, and it detects any faults at an early stage and prevents loss. These methods

being data-driven, they can diagnose faults in any machinery independent of the complexity in configuration of gearbox. These properties of data-driven approaches make them widely applicable and less complex to implement as opposed to using only signal processing methods. The proposed methodology is explained in the next Chapter.

Chapter 3

Proposed Methodology and Numerical Simulations

In this Chapter, development of the automated fault diagnosis algorithm is explained and illustrated using a simulated signal. The overall steps undertaken are first summarized in an itemized form, followed by a detailed explanation of the procedures used.

3.1 Proposed Methodology

The basic approach carried out in this thesis is as follows.

1. Vibration signals contain noise when collected from a gearbox, so the signals have to be filtered in order to discern and work with the desired ranges of frequencies. In this research, an empirical wavelet decomposition (EWD) approach is used for filtering the vibration signals (see Appendix A).

2. The next step is to transform the raw signal into a condition-indicator space. The condition indicators derived from the raw signal here are: maximum, minimum, mean, standard deviation, root-mean-square, skewness, kurtosis, normalized sixth moment, crest factor, amplitude square, pulse factor, root amplitude, margin factor and operating energy. Expressions for the condition indicators are shown in Table 3.1. Based on various studies conducted in this thesis, at least 1000 windows of the signal (which result in 1000 condition indicators) are required for statistically meaningful results. PCA is performed on this feature space and first two components which contain the maximum variation are used for next step (see Appendix C).

Table 3.1: Condition Indicators

No.	Condition indicator	Description	Expression
1	MX	Maximum	$x_{mx} = max(\|x_i\|)$
2	MN	Minimum	$x_{mn} = min(\|x_i\|)$
3	ME	Mean	$x_{me} = \frac{\sum x_i}{n}$
4	SD	Standard deviation	$x_{sd} = (\frac{1}{n-1}\sum_{i=1}^{n}(x_i - x_{me})^2)^{1/2}$
5	RM	Root mean square	$x_{rm} = (\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{1/2}$
6	SK	Skewness	$x_{sk} = \frac{\sum_{i=1}^{n}(x_i - x_{me})^3}{(n-1)x_{sd}^3}$
7	KT	Kurtosis	$x_{kt} = \frac{\sum_{i=1}^{n}(x_i - x_{me})^4}{(n-1)x_{sd}^4}$
8	NS	Normalized sixth moment	$x_{ns} = \frac{\sum_{i=1}^{n}(x_i - x_{me})^6}{(n-1)x_{sd}^6}$
9	CF	Crest factor	$x_{cf} = \frac{x_{mx}}{x_{rm}}$
10	AS	Amplitude square	$x_{as} = \sum_{i=1}^{n} x_i^2$
11	PF	Pulse factor	$x_{pf} = \frac{x_{mx}}{x_{me}}$
12	RA	Root amplitude	$x_{ra} = (\frac{1}{n}\sum_{i=1}^{n} \|x_i\|^{1/2})^2$
13	MF	Margin factor	$x_{mf} = \frac{x_{mx}}{x_{ra}}$

3. Then, a Gaussian mixture model (GMM) (Appendix E) clustering based on expectation maximization (EM) algorithm (Appendix F) is undertaken. This step calculates the sufficient parameters of a two-dimensional Gaussian in PC space to describe the probability density of any cluster.

4. Once the condition indicators and the GMM models are calculated, the logarithm of Mahalanobis distance (l) (see Appendix D) of each point (CI) from the centroid of first cluster is calculated using the mean and variance of the cluster. l is then used as the process variable for monitoring using SPC and monitored until SPC detects a first change.

5. As the number of operating conditions is assumed for a machine, there will be a point when the number of clusters remains constant. Once this is satisfied for a period of time (this is determined from known operating conditions), incoming data is then classified as belonging to a particular cluster. This process is carried out by calculating the l from any new data point (a 2 dimensional vector) to the respective centroids of the identified clusters. The minimum value of l (l is a vector of size c where c is the number of clusters) determines the cluster a given data point belongs to. For the case of novelty detection in this thesis, $c = 1$, and the threshold for the cluster is determined as a percentile value (3σ) while performing SPC.

In its initial implementation, the calculation of condition indicators is automated in a NI CompactRio® system. Following this, the calculation of GMM parameters and SPC are performed offline. This approach promises to be efficient in industrial applications for automated fault detection. In the subsequent Chapters, this algorithm is implemented on a set of simulated and experimentally acquired signals and key results are presented.

3.2 Simulated Vibration Signals from Gearbox

For the sake of illustration, a gear in its pristine state can be represented by a pure sinusoid [21] whose central frequency matches with the meshing frequency, i.e., the shaft rotation frequency multiplied by the number of gear teeth. Meshing defects in gear are manifested through the appearance of sidebands around the meshing harmonic, which can typically be represented by amplitude modulating and frequency modulating (AM-FM) signals [51].

Consider an example constituted from harmonics with 3 AM-FM components representing progressive degradation, with additive white noise. This model can be written as:

$$s(t) = \sum_{k=1}^{n_H} A_1 \sin 2\pi k f_G t \; + \sum_{k=k_1,k_2,\ldots k_{sb}} (1 + A_2 \sin 2\pi f_2 t)\,(\sin 2\pi k f_G t + A_3 \sin\,(2\pi k \beta t)) \tag{3.1}$$

In Equation 3.1, the first term represents a gear signal in its pristine state. n_H is the number of harmonics and f_G is the gear-meshing frequency. The second term represents the AM-FM components —or the sidebands —which appear at specified locations k_1, k_2,...etc. k_i is the i^{th} harmonic of the gear-meshing frequency corresponding to which the AM-FM component is added. The quantities A_1, A_2, A_3 are constants, which are user defined, and related the amplitudes of the sinusoidal with the AM-FM parts. β is a constant, which lies between 0 and 1.

Using the above notation, consider a signal with 12 harmonics. The fundamental gear meshing frequency is assumed to be 80 Hz (i.e. $f_G = 80$) and the AM-FM components added are 4^{th}, 8^{th} and 12^{th} harmonics, i.e., at frequencies of 320, 640 and 960 Hz, respectively. The sampling frequency is 2048 Hz, β is 0.5 and f_2 is 2 Hz. The 3 windows of signals (each window is 20 s long) in which the first window comprises of the pristine

component in Equation 3.1 and 1 AM-FM component (i.e. $k_1 = 4$), the second window with 2 AM-FM components, and the third window with 3 AM-FM components alongside the pristine component (refer Figure 3.1). The standard deviation of the clean signal is 8.8 Hz and that of the added noise is 5 Hz. The signal from gearbox is a concatenation of the 3 windows and is shown in Figure 3.2.

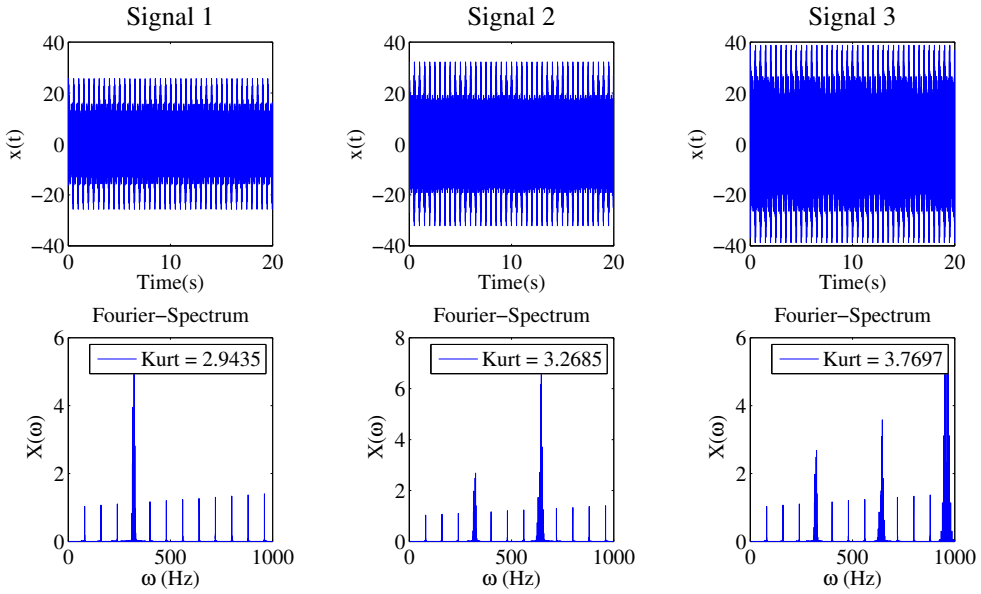


Figure 3.1: Three Segments of the Gear Signal

Table 3.2: Calculation of GMFs

	n_t	ω (RPM)	f (Hz)	G_i GMF (Hz)
Driving pinion	23	1738	29	667
Driving wheel	37	1080	18	667
Worm shaft	5	1080	18	90
Output gear	27	200	3.33	90

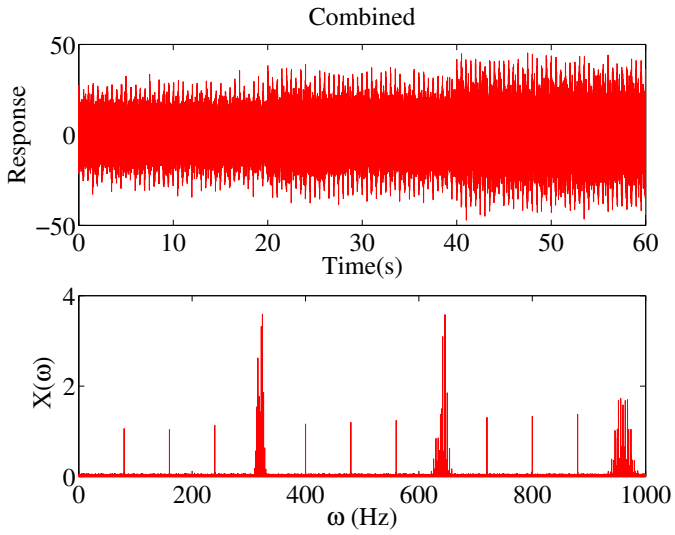


Figure 3.2: Concatenated Signal for Analysis

3.3 Condition Indicators (Features)

From the signal in Section 3.2, the condition indicators (features) are calculated. The signal $s(t)$ is of length 122880 and is divided into windows of size 1000 each. The features described in Table 3.1 are calculated for each window where x_i is the i^{th} sample of $s(t)$. A feature vector of size $Y_{m\times n}$ is obtained where $m = 122$ and $n = 13$. The former is the number of partitioned windows and the latter is the size of condition indicators used. The plots of each of $Y_{m\times l}$, where $l = 1, ..., n$, are shown in Figures 3.3, 3.4, 3.5 and 3.6.

Amongst the features shown, clearly three states are observed in the standard deviation, RMS, amplitude square and root amplitude. Maximum and minimum show slight similar behaviour, while kurtosis and normalized sixth moment do not show any distinction between the first and second states, while the third state is noticeable. The remaining condition indicators do not show any noticeable trend.

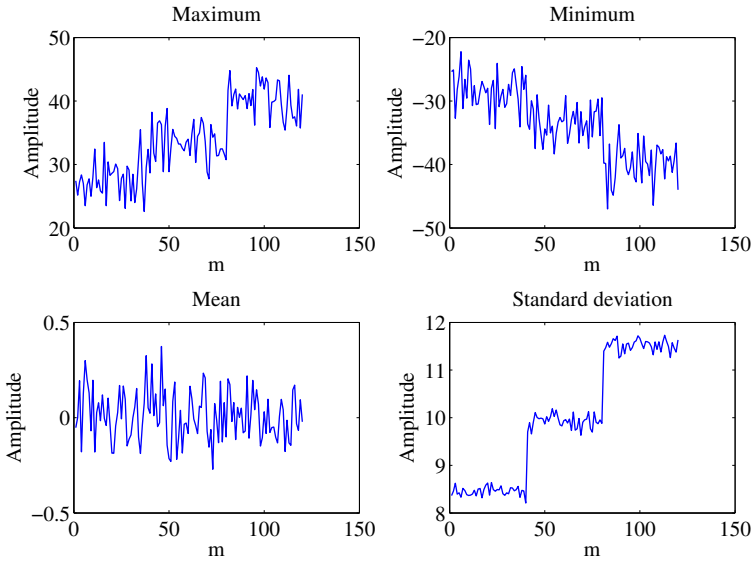


Figure 3.3: Features Set 1

Perhaps of particular importance for this study, not all the 13 features reflect a common trend in the condition of the gearbox and moreover working with 13 dimensional feature space is computationally expensive. Hence, it is useful to reduce the dimensions of these features, which is accomplished using PCA (see Appendix C). The principle behind PCA is a linear transformation of the feature space during which the dimensionality is preserved while information is reorganized. Let μ_Y and Σ_Y be mean and covariance of Y, respectively. The feature vector, Y is standardized using Equation 3.2:

$$Y_{Std} = (Y - \mu_Y) \times diag(\Sigma_Y)^{-1/2}. \tag{3.2}$$

After performing singular value decomposition (SVD) on $\Sigma_{Y_{Std}}$ (the covariance matrix of Y_{Std}), the transformed feature vector T and Σ_T are derived. The relation between $\Sigma_{Y_{Std}}$,

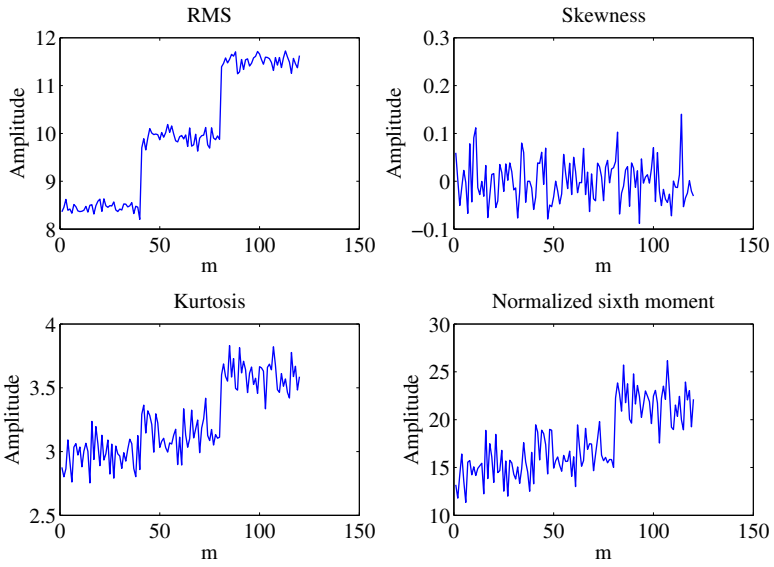


Figure 3.4: Features Set 2

T and Σ_T is given by:

$$\Sigma_{Y_{Std}} = T \times \Sigma_T \times T'. \tag{3.3}$$

The transformed feature vector X is given by:

$$X = Y_{Std} \times T. \tag{3.4}$$

PCA results in a transformation vector where the components are in the decreasing order of their rank of their eigenvalues, which explain the variance in the components. For fault diagnosis, the first two components (columns) of X are used to represent the crucial information content of vibration data (which explain most of the variance in the data), which are hereafter referred to as PC_1 and PC_2.

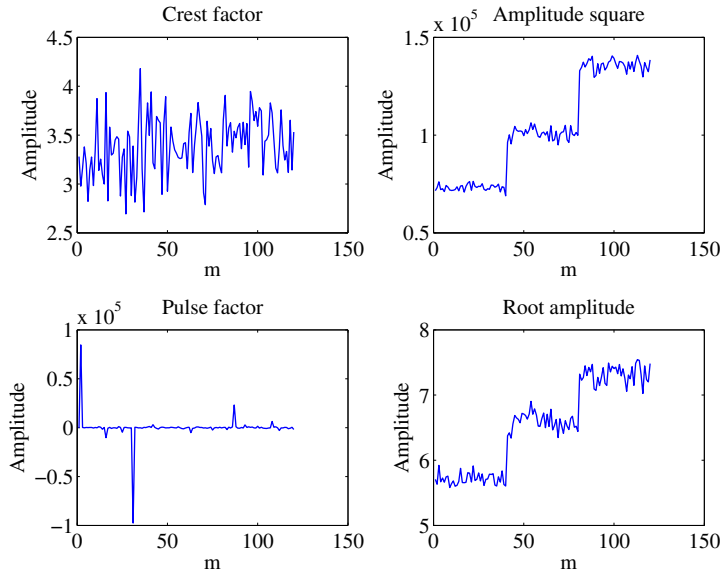


Figure 3.5: Features Set 3

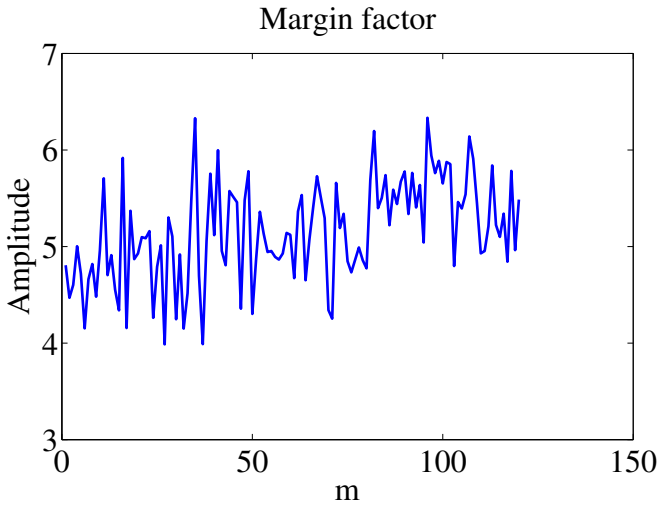


Figure 3.6: Features Set 4

3.4 Fault Diagnosis

The basic premise of the approach pursued here is that vibration data contains fault information; hence, analyzing the features derived from vibration data allows us to discern the condition of the machine. This starts with analyzing the PCs with the aid of a scatter plot (Figure 3.7). This is premised on the assumption that the data is well separable, i.e., there is a clear distinction between the three states of fault conditions in the simulated signal.

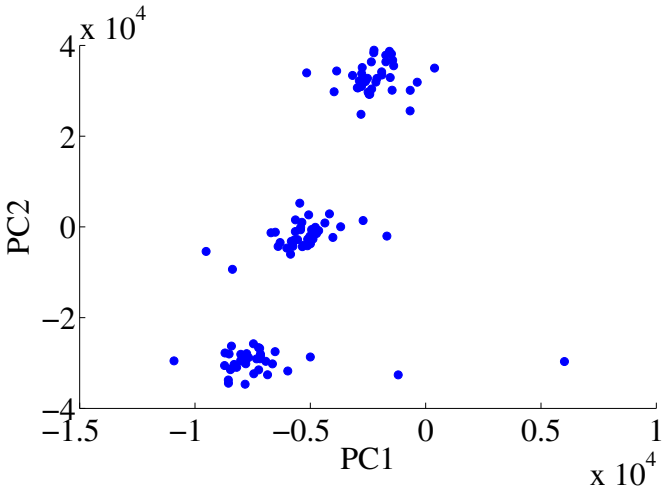


Figure 3.7: Components $PC1$ and $PC2$

The next step is to cluster the data and find the parameters of the clusters that will be used for novelty detection. In this work, GMM is used for clustering the data. GMM is a weighted sum of Gaussian component densities given by the equation:

$$p(X|\lambda) = \sum_{i=1}^{M} w_i g(X|\mu_i, \Sigma_i) \tag{3.5}$$

where X is the n dimensional feature vector, M is the number of components (clusters)

in X, $w_i, i = 1, ..., M$ are mixture weights and λ is used to represent the parameters, $\lambda = \{w_i, \mu_i, \Sigma_i\}, i = 1, ..., M$. $g(X|\mu_i, \Sigma_i), i = 1, ..., M$ are the component Gaussian densities and each component density is a n-variate Gaussian function of the form

$$g(X|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} exp\left\{-\frac{1}{2}(X - \mu_i)^{'}\Sigma_i^{-1}(X - \mu_i)\right\} \tag{3.6}$$

where, μ_i is the mean vector and Σ_i is the covariance matrix. w_i satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$.

The parameters of the Gaussian densities are determined from X using maximum likelihood (ML) estimation, such that the likelihood of observing X is maximized. This is explained next.

First, a likelihood function is defined as,

$$p(X|\lambda) = \prod_{t=1}^{T} p(X_t|\lambda) \tag{3.7}$$

Since the above expression is a non-linear function of λ and direct maximization is not possible, ML parameters are estimated iteratively using the expectation-maximization (EM) algorithm [17] (see Appendix F).

The main difficulty in implementing ML estimation using EM algorithm is converging to a local optimum. To overcome this, the initial values of μ_i and Σ_i and the number of clusters are obtained using the k-means objective (see Appendix G) for clustering the data, which helps in faster convergence [40]. Figure 3.8, shows the centroids of the clusters obtained after performing k-means iteration. The number of clusters is validated using the elbow principle [57], according to which the variance of the clusters (within-cluster sum of squares, a scalar value for each k) at each k (assumed number of clusters in each iteration) is calculated. The value of k for which there is a first significant slope change is the optimal

k. From Figure 3.9, for this problem there are 3 clusters in the data and accordingly the k-means objective results in centroids for the 3 clusters.

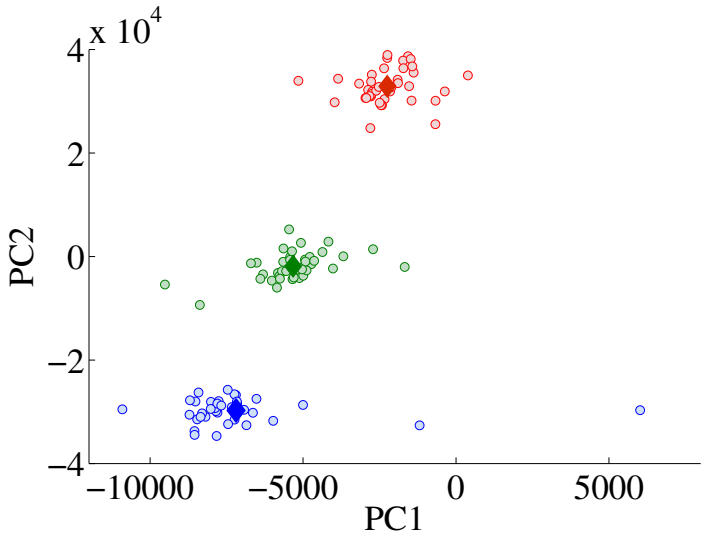


Figure 3.8: k-means Clustering

Using the k-means procedure, μ_i and Σ_i of X are calculated and are used to initialize the EM algorithm. The objective of EM algorithm is to estimate a new $\bar{\lambda}$ such that $p(X|\bar{\lambda}) \geq p(X|\lambda)$ where $p(X|\lambda)$ is obtained from Equation 3.7. The refining iterative approach is explained in Appendix F. The final results obtained through this procedure are shown in Figure 3.10.

The *a posteriori* probability (also known as membership score) is the belongingness of a datapoint t to cluster i, given by Equation 3.8 and the Figure 3.11 shows the membership score.

$$Pr(i|X_t, \lambda) = \frac{w_i g(X_i|\mu_i, \Sigma_i)}{\sum_{k=1}^{M} w_k g(X_t|\mu_k, \Sigma_k)} \tag{3.8}$$

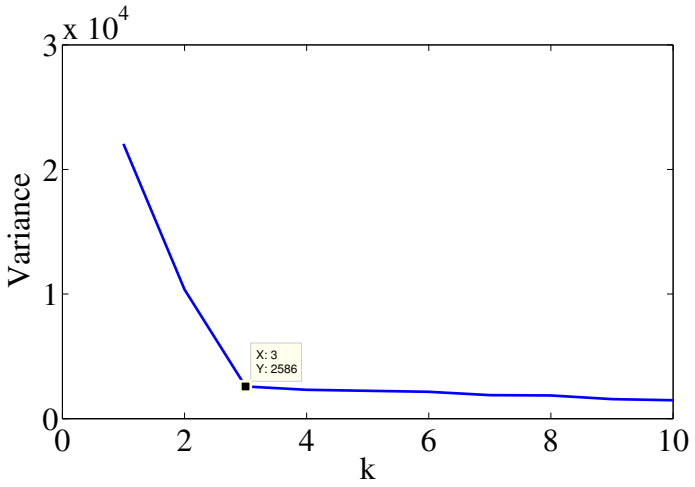


Figure 3.9: Number of Clusters

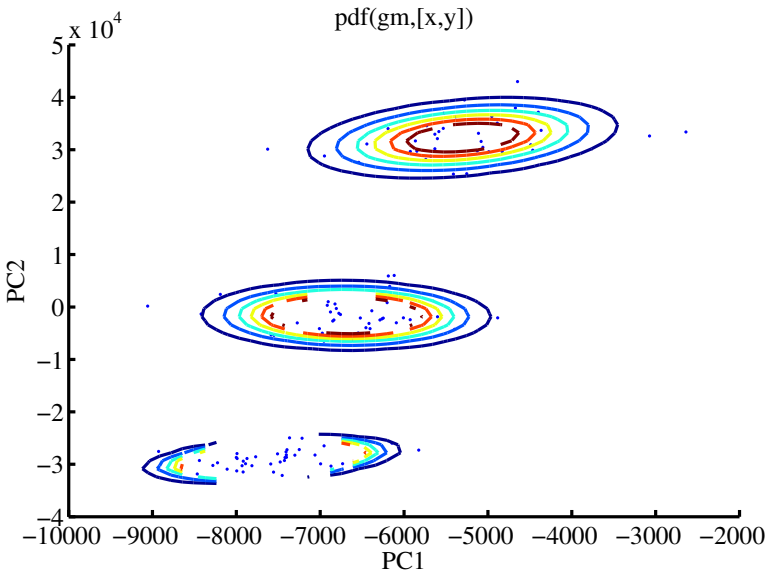


Figure 3.10: GMM Clustering

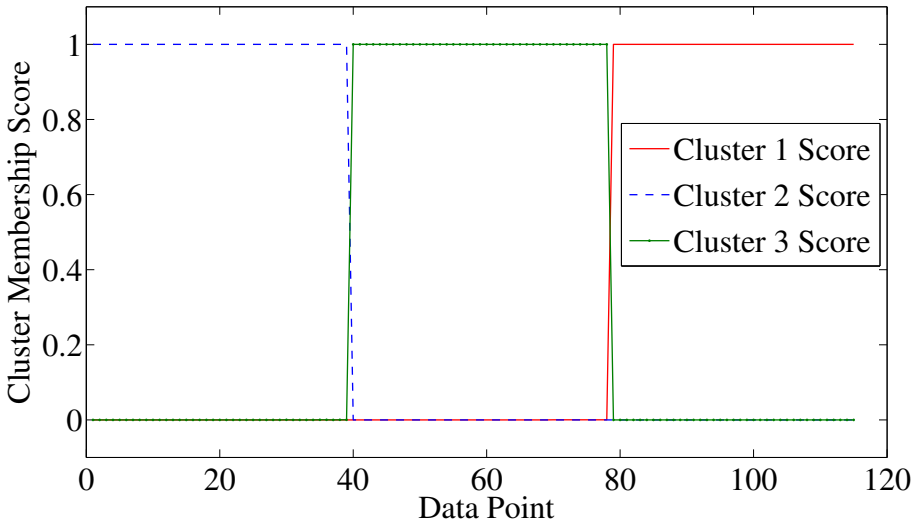


Figure 3.11: Membership Score

Once the parameters of the clusters are obtained from GMM, the fault conditions are described next with the aim of automating the fault detection process. To this end, the process variables for SPC are defined in the next Section along with the automation process.

3.5 Automation

In the previous Section, because there is more than one cluster estimated through GMM, it is inferred that the simulated signal represents multiple faulty states. Additionally, it is important that the fault is identified well in advance to schedule preventive maintenance. This process is implemented within the framework of novelty detection to provide early warning when gearbox is about to fail, and the necessary actions can be contemplated. This process of automated fault diagnosis is explained in the following Section.

3.5.1 Novelty Detection

Novelty detection is the process of identifying a *novel* condition, which is synonymous with a fault for the purposes of this study. A baseline is first established when the gearbox is in its healthy or *pristine* state. This does not necessarily mean that the machine is new, only that the data is acquired when it is known that the machine is in satisfactory working condition.

The baseline is established as follows. An arbitrary number of points (say, at data point with index $t = 30$) are collected first. Parameters μ_t and Σ_t for the GMM clustering procedure, as described earlier, are calculated using the feature set X_t (denotes feature set calculated using the data obtained until time t). It is assumed that all the data in X_t belongs to the pristine condition of the gearbox, and μ_t and Σ_t are used as parameters for the pristine model.

As new data is acquired, this is compared against the data from the pristine condition and the level of deviation from its normal condition has to be quantized as a novelty score. In order to determine this novelty score, MD is calculated between each incoming point X_l and μ_t using the covariance matrix Σ_t, the equation is given by:

$$MD(l) = \sqrt{(X(l) - \mu_t)' \Sigma_t^{-1} (X(l) - \mu_t)} \tag{3.9}$$

The logarithm of the MD is then used as novelty score (η). η_l denotes the novelty score of data at index l where $l = t, ..., M$. It is assumed that η follows normal distribution and a threshold is set using μ_η and σ_η.

Figure 3.12 shows variation of η with index of data points. Dotted lines show the threshold set at $\pm 3\sigma_\eta$. It is observed that a novelty is detected in η at $l = 40$. So there are two fault conditions observed in the vibration signal which is consistent with the simulated

signal (see Section 3.2). For automation purposes, the system should alert incipient fault at $l = 40$, which is implemented using SPC as described in the next Section.

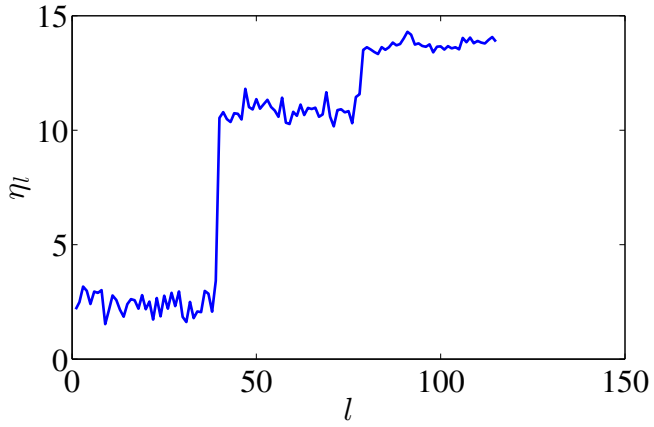


Figure 3.12: Novelty Score

3.5.2 Statistical Process Control

SPC enables us to define monitoring and setting control rules (values) when the system detects an incipient fault. In order to define this process variable, the novelty score η is used. Exponentially weighted moving-average (EWMA) chart is applied to monitor the process. EWMA (denoted by z_l at time l) is an adaptive mean value which forgets previous values at exponential rate. The expression for z_l is given by,

$$z_l = \beta\eta_l + (1-\beta)z_{l-1} \tag{3.10}$$

which can also be written as

$$z_l = \beta\sum_{i=0}^{l-1}(1-\beta)^i\eta_{l-i} + (1-\beta)^l z_0 \tag{3.11}$$

where β is the *forgetting factor*, which lies between (0,1); usually $\beta = 1$ offers an advantage when estimating the value [22]. While η_l is independent with the mean value μ_{η_l}, the mean for z_l is given by

$$E(z_l) = \mu_{\eta_l}(1 - (1 - \beta)^l) + (1 - \beta)^l z_0 \tag{3.12}$$

and because $\beta \in (0, 1)$ and $l \to large$, $E(z_l)$ reduces to

$$E(z_l) = \mu_{\eta_l} \tag{3.13}$$

The standard deviation σ_{z_l} (rather, the variance of z_l) is defined by [46],

$$\sigma^2_{z_l} = \sigma^2_{\eta_l} \left(\frac{\beta}{2 - \beta} \right) [1 - (1 - \beta)^{2l}] \tag{3.14}$$

and the lower and upper control limits are defined by [46],

$$UCL, LCL = \mu_{\eta_l} \pm \theta \sigma_{\eta_l} \sqrt{\frac{\beta}{2 - \beta}[1 - (1 - \beta)^{2l}]} \tag{3.15}$$

where θ is the design parameter and controls the sensitivity of the control chart.

For the numerical example under study, the control rules for the process variable η_l are set using $\theta = 3$. This means that the system is said to be deviating when all 3 of any 3 successive η_l lie above or below the UCL/LCL. Plotting the EWMA control chart from the data point $t = 30$ as shown in Figure 3.13, it is observed that at $l = 39$, the process variable deviates the control rules. Therefore, at $l = 39$ the SPC algorithm detects that the signal is behaving abnormally and the system has to be monitored.

Progressing further, forgetting the control variable and control limits prior to $l = 39$, a new window for monitoring is considered and the process variable and control limits are calculated using the definitions above. At $l = 40$, the system triggers a fault condition (see Figure 3.14).

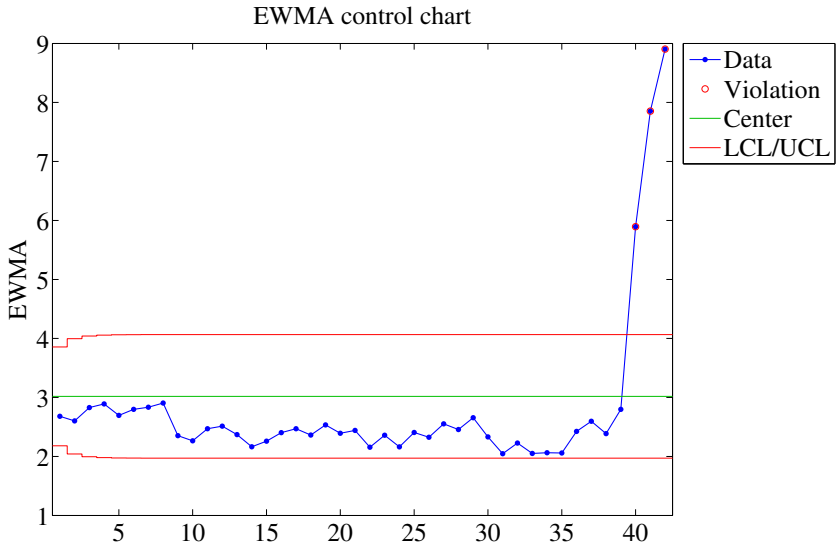


Figure 3.13: Monitoring Process - Until the First Alert

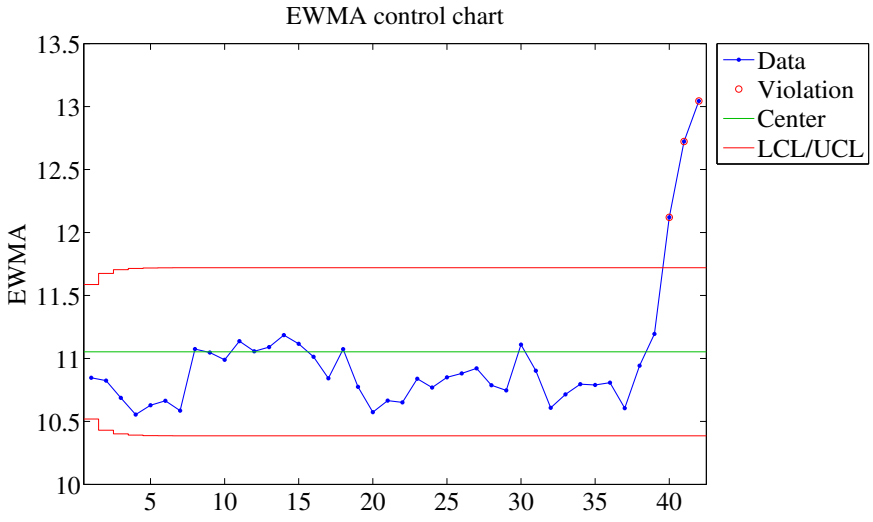


Figure 3.14: Monitoring Process - After the First Alert

3.6 Summary

The proposed methodology is described in this Chapter with application to a simulated signal. The results are consistent with the simulated faults and show that there are 3 clusters in the vibration data and the GMM algorithm results in the correct number of clusters and their parameters. It is observed at this stage that the proposed method is sensitive to faults in the gearbox and the sensitivity is improved further by using a novelty score. In conjunction with SPC, this approach offers a viable condition monitoring scheme where it is essential that the the system in place is automated and can detect incipient faults. This methodology is implemented on data from an experimental laboratory set-up in the next Chapter and the results are validated.

3.7 Limitations of the Proposed Approach

Although most limitations in existing work (see Section 2.3) are addressed in this work, there are few points of improvement in the method developed. These are listed below:

1. The proposed approach uses parametric method of GMM, which is based on EM for parameter estimation. Convergence of EM to a local optimum depends on number of iterations and initialization.

2. The study has been performed on vibration data collected at a single RPM, which is subject to vary in real time.

3. SPC has been sensitive to alert when a change is detected. In a single RPM case, the change is mapped to a faulty condition, but in case of a varying RPM, the change

can correspond to change in RPM. The method has to be modified to identify this change as a potential fault or a change in RPM.

Chapter 4

Laboratory Experiments

While the framework and the procedures have been described in the previous Chapter, these have to be implemented on real systems prior to field implementation. In this Chapter, vibration data from controlled experiments on an experimental set-up that simulates real world machinery, is utilized to demonstrate the applicability of the methodology. Specifically, a drivetrain diagnostics simulator (DDS) is used for simulating faults in rotating machinery. This Chapter explains the DDS, its functionality, and the results obtained using the proposed methodology.

4.1 Drivetrain Diagnostics Simulator

4.1.1 Configuration and Details

The DDS is designed by SpectraQuest to simulate industrial drivetrains, especially as an experimental research tool. Figure 4.1 shows the drivetrain simulator, which is designed

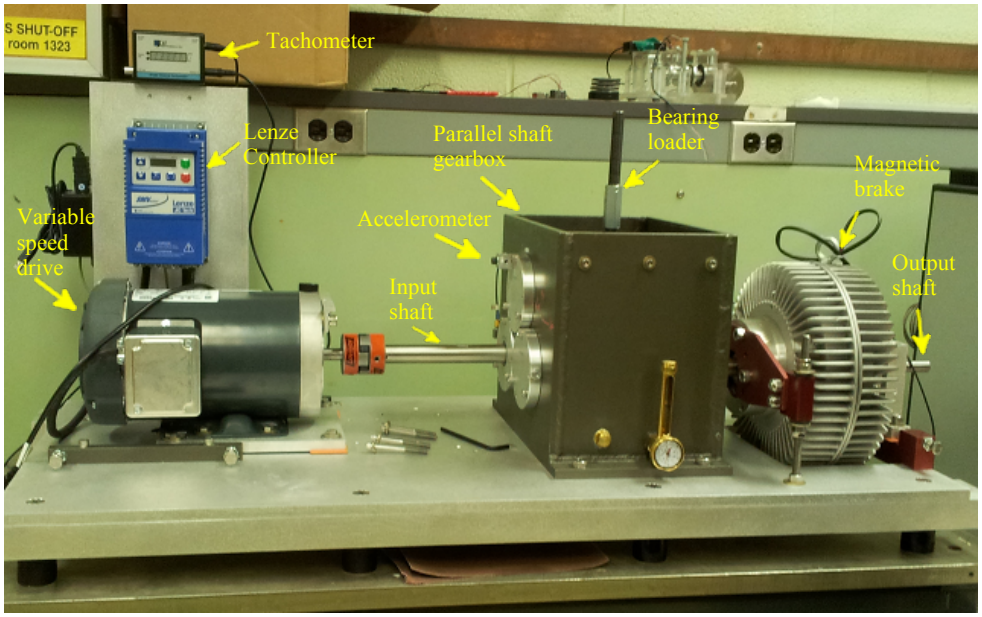


Figure 4.1: Drivetrain Diagnostics Simulator (DDS)

for studying common gear and bearing faults. It weighs about 96 kg and has a modular design, which enables easy reconfiguration and simple rules of operation, making it ideal for research projects. Its components are instrumented to high tolerances and avoids conflicting vibrations during its operation. It consists of a 2 stage oil-lubricated parallel shaft gearbox with rolling bearings, a bearing loader, and a programmable magnetic brake. The elements of the DDS are designed such that a large number of configurations of the drivetrain can be achieved and used for experiments for condition monitoring based on vibration analysis, lubricant analysis, and wear particle analysis. It is designed to handle heavy loads and it is wide enough for gear replacement and to accommodate set up, installation, and monitoring devices. The gears can be configured to increase or decrease gear ratio.

Given the modular setup of the DDS, it is designed to conduct tests by replacing the pristine gears with faulty ones. A 3 HP variable frequency AC drive with a programmable Lenze controller allows adjusting the frequency of the input shaft (see Figure 4.1). A built-in tachometer measures the rotation speed of the shaft to measure the transmission error. The accelerometers are threaded to mounting disks, which are rigidly attached to the surface of the gearbox with the aid of screws.

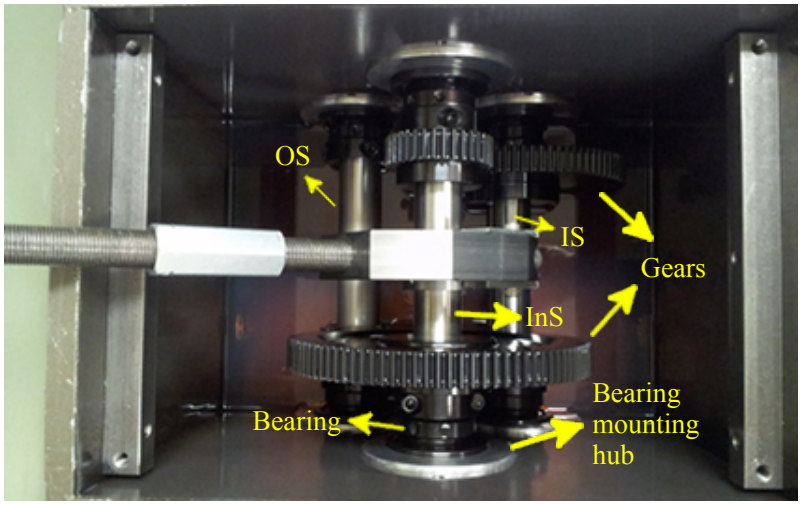


Figure 4.2: Two Stage Parallel Shaft Gearbox - Topview. IS: Input Shaft, InS: Intermediate Shaft, OS: Output Shaft

The parallel shaft gearbox has a glass top for visibility during its operation. Figure 4.2 shows the top view of the parallel shaft gearbox. There are three shafts - an input shaft, an intermediate shaft, and an output shaft. Gears can slide along the shafts to alter system stiffness and allow for mounting additional devices. The shafts are connected to the walls of the gearbox with bearings and bearing mounting hubs for smooth movement. The bearing mounting hub of the intermediate shaft allows for adjusting the clearance (eccentricity shown in Figure 4.3). This allowance is provided to adjust the clearance

to study backlash and its consequences when the gearbox is in operation. Intentionally damaged or worn out gears and bearings can be placed on the shafts to study their effects on vibration signature. Roller bearings used in the present study make allowance for axial movement of the shafts by an adjustable spring mechanism provided by Belleville spring washers placed at the output end of the shafts. A computer controlled magnetic brake is connected directly to output shaft to provide loading. When loading is increased, the amplitude of the vibration signals are commensurately higher, making it easy to discern the fault conditions in the gearbox.

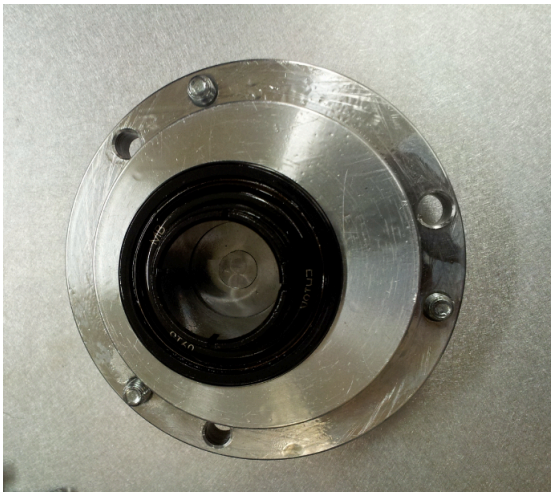


Figure 4.3: Eccentric Mounting Hub for Studying Backlash

The modular design of the DDS allows for easy replacement of components depending on the objective of study. The motor and variable speed drive facilitate tests to be performed at different RPM of the input shaft and the tachometer measures the frequency of rotation of the input shaft. The parallel shaft gearbox is an important component of the DDS and it allows for various types of studies for research. The gears and bearings can be replaced depending on the fault diagnostics objective. The loading mechanism can be

modified to radial or torsional to study the respective effects on vibration signals from the gearbox. The shafts are designed to accommodate spur or helical gears and roller or sleeve bearings for rotation of the shaft. The design of the gearbox provides abundant space to choose the mounting location for multiple accelerometers. Gear faults can be studied in detail by replacing healthy gears with chipped, missing, and cracked teeth gears.

4.1.2 Experimental Set-up

To illustrate the fault detection algorithm using DDS, experiments involving faulty gears and bearings are conducted. As mentioned earlier, DDS allows for experiments using prefabricated faulted gears and bearings. Hence, several experiments are designed such that the vibration data simulated encompasses major fault conditions that are commonly found in gears and bearings.

Defective Gears

Faults in gears are caused due to improper meshing between coupling units, or wear over time. Figure 4.4 shows faults that are commonly observed in gears. Figure 4.4a shows a gear with chipped tooth and Figure 4.4b shows a gear with missing tooth. These gear faults cause variation in the vibration signals which are measured using accelerometers. Another type of fault commonly observed in gears is a root crack, shown in Figure 4.4c. In this work, data from a chipped tooth gear and a missing tooth gear is collected using DDS, and the analysis described earlier is performed.

(a) Chipped Tooth (b) Missing Tooth (c) Root Crack

Figure 4.4: Gear Faults

Bearings

Bearings are used to constrain movement between parts of the gearbox to desired axis, while providing smooth movement between components. In the DDS, bearings are installed between the shafts and the mounting hubs (see Figure 4.5). A sectional view of the bearing is shown in Figure 4.6. It consists of three main components - (i) inner race, i.e., the inner ring that is connected to the shaft, (ii) outer race and (iii) the rollers (usually spherical or cylindrical) between the inner race and the outer race. Inner race rotates at the RPM of the shaft.

Faults in bearings are generated due to fatigue, wear, improper installation and lubrication, or due to manufacturing defects. Although the bearing faults are not visible as long as the bearing is in working condition (unlike faults on gears), upon dismantling three types of faults have been observed in bearings: a ball fault, fault in inner race, and fault in the outer race (see Figure 4.7). A crack or chipping in the outer race is called an outer race fault and such a fault in the inner race is called an inner race fault. There could be brinnel marks and indentations on both outer and inner races due to excessive loading,

Figure 4.5: Bearing

or misalignment in the outer and/or inner races. Spalling due to fatigue and fracture of running materials in the inner race, outer race and ball creates unfamiliar frequencies in the vibration signals from the bearings. In this research, bearings with induced faults are used in some of the test cases. Two types of faulty bearings are used - inner race fault and outer race fault.

4.1.3 Replacement Procedure

For gear diagnostics using DDS, the gears are replaced by their faulty counterparts. The top of the gearbox is opened and the necessary parts are removed in a sequence. By observing the gearbox, for any replacement, the output side is shielded by the magnetic brake because of which all the replacement is carried out on the input side of the gearbox. Figure 4.8 illustrates gear ratios of the four gears connected to the three shafts. The smaller

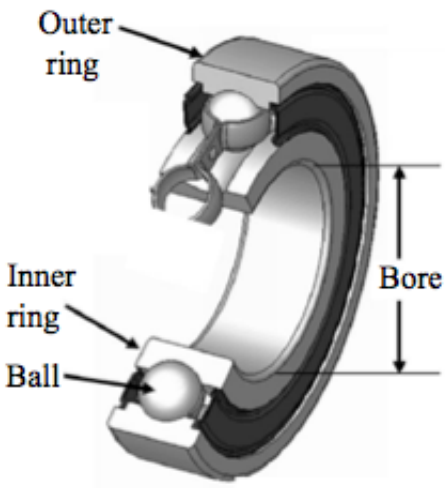


Figure 4.6: Bearing - Section

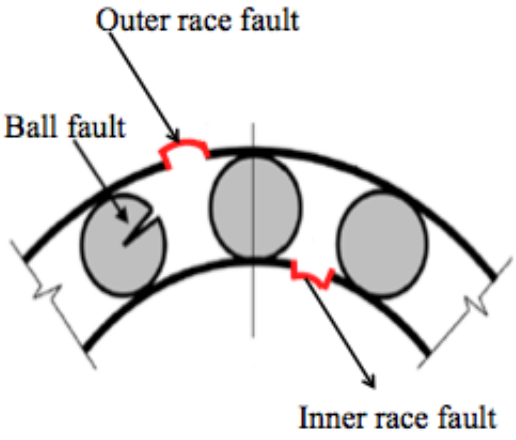


Figure 4.7: Bearing Faults

gear on the intermediate shaft is connected to the output side of the gearbox. The gear fault kit provided by the manufacturer includes faulty specimens for the gear with 36 teeth on the output side of the intermediate shaft (Figure 4.8) only.

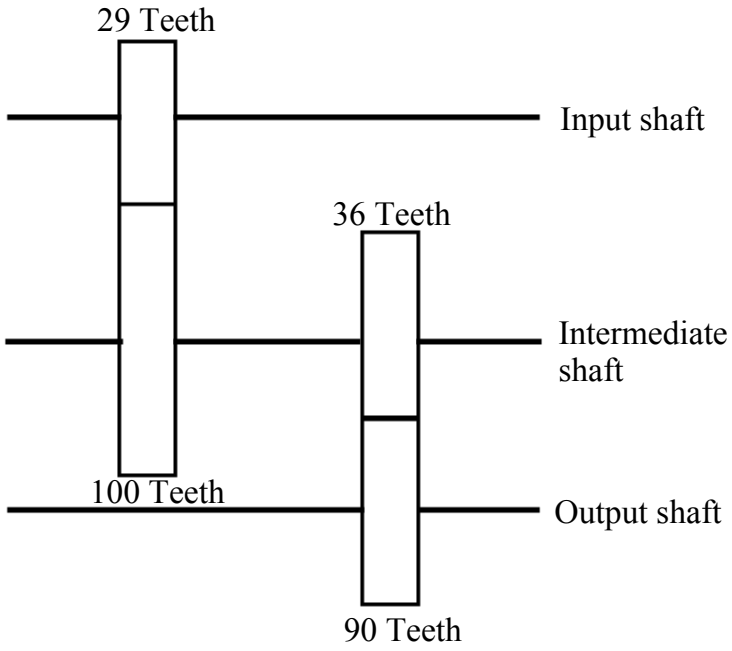


Figure 4.8: Gears Meshing

First, the bearing mounting hub on the input side of the intermediate shaft is removed. The steps are illustrated in the Figure 4.9. The three screws on the hub are removed to loosen the mounting hub (Figure 4.9a). The screws connecting the input side gear are loosened and the gear is slid towards the output side along the intermediate shaft (Figures 4.9b and 4.9c). The bearing is loosened on the input side and also on the output side of the intermediate shaft (Figures 4.9d and 4.9e). The hub is removed either by pushing from

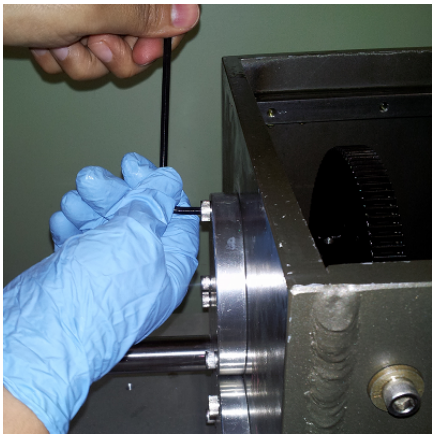
the inside or with the help of pusher screws provided for the purpose, as shown in Figure 4.9f. The Belleville spring washers can be seen in the hub after removing the gears from the gearbox.

After removing the mounting hub (Figure 4.10), the intermediate shaft is removed. As shown in Figure 4.11, the larger gear is on the input side and the smaller gear is on the output side. To replace the smaller gear, the intermediate shaft is lifted from its position and removed from above (see Figure 4.12). The position of the gear is noted and removed from the shaft. The desired defective gear is screwed on to the shaft at the correct position and the shaft is replaced. The bearing screws are then tightened on the output side of the shaft and the mounting hub on the input side is placed in its position by aligning the position of the bore hole and is pushed on to the wall of the gearbox. The gear and bearing screws on the input and output sides of the shaft are tightened.

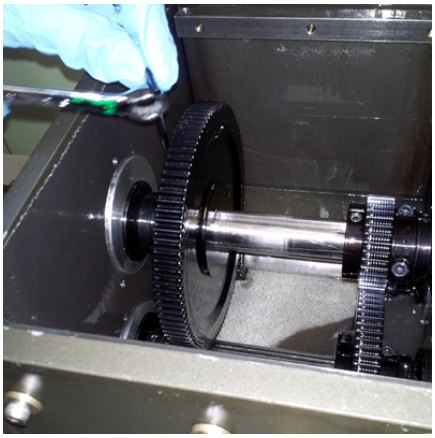
4.1.4 Data Collection

For all gear configurations, vibration signals are measured by fixing accelerometers at the desired positions. A Dytran 3263A2 model triaxial accelerometer is used to collect data in three directions. Figure 4.13 shows the accelerometer set-up mounted to one of the hubs of the intermediate shaft. The accelerometer is screwed onto a mounting disk (Figure 4.13a) and the mounting disk is fixed to the surface (Figure 4.13b). In order to start the DDS, the Lenze controller is turned on. Figure 4.14 shows a close-up view of the controller. The frequency of the input shaft is adjusted on the controller using the user panel. The tachometer located above the controller displays the shaft rotation frequency. Vibration signals are measured at speeds of 23, 26, 29, 32, 35, 38, 41, 44, 47 and 50 Hz.

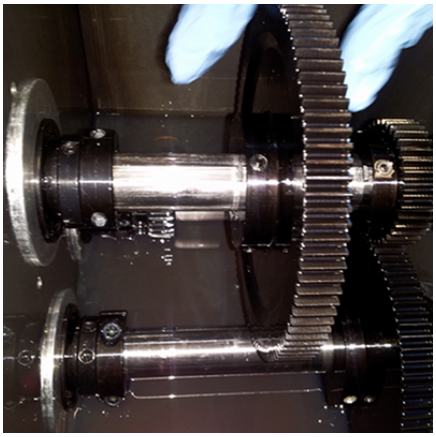
Two sets of tests are carried out using the described DDS set-up. In the first set, tests



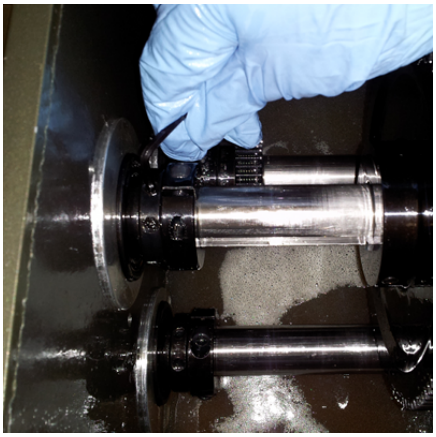
(a) Loosen Mounting Hub



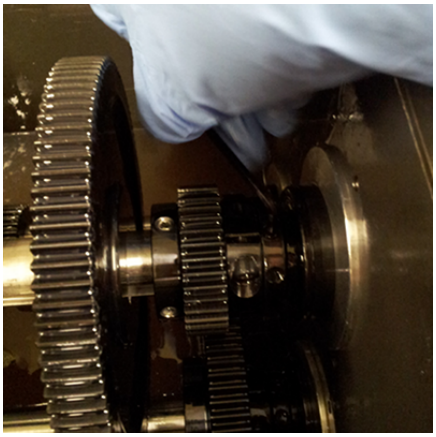
(b) Loosen Gear Screws



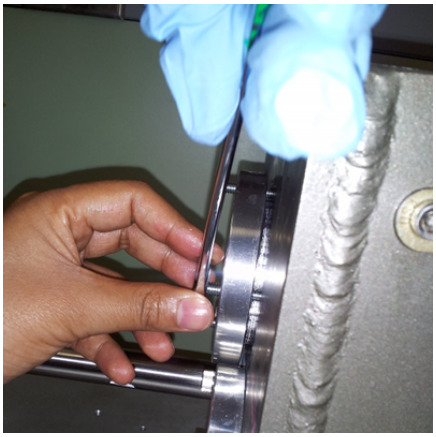
(c) Slide the Gear



(d) Loosen Bearing Screws



(e) Loosen Output Bearing



(f) Remove Hub

Figure 4.9: Steps to Remove Intermediate Shaft

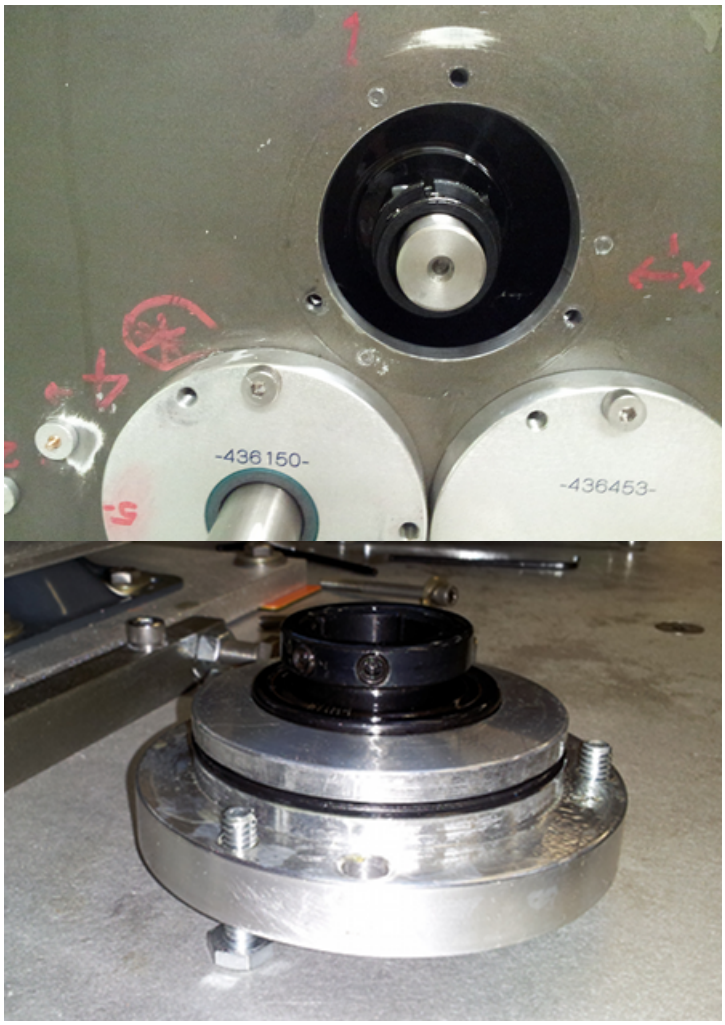


Figure 4.10: Bearing Mounting Hub

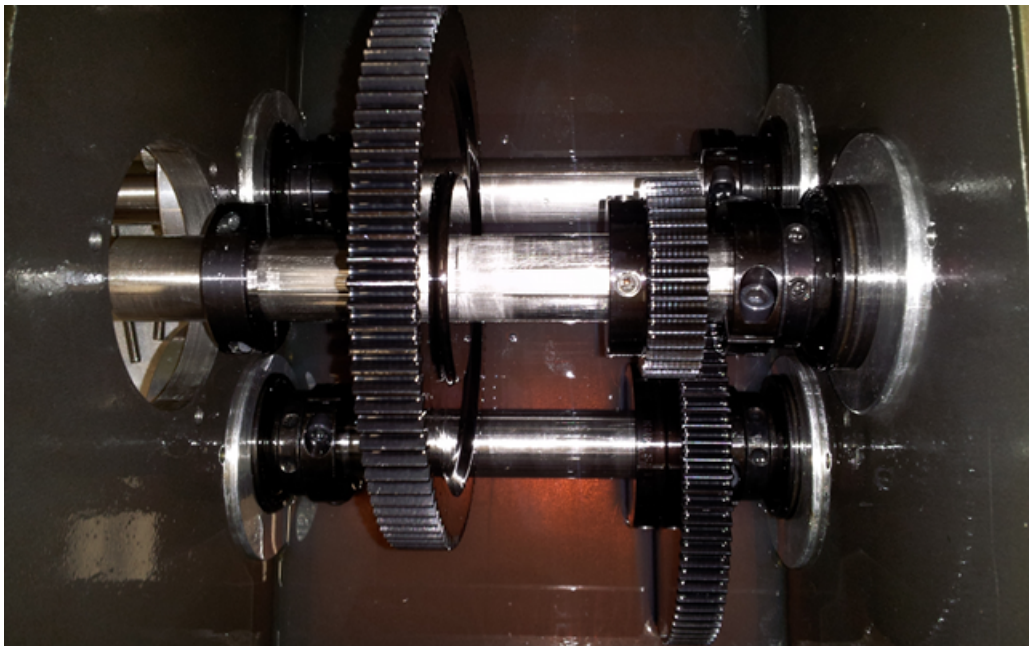


Figure 4.11: Removal of Intermediate Shaft

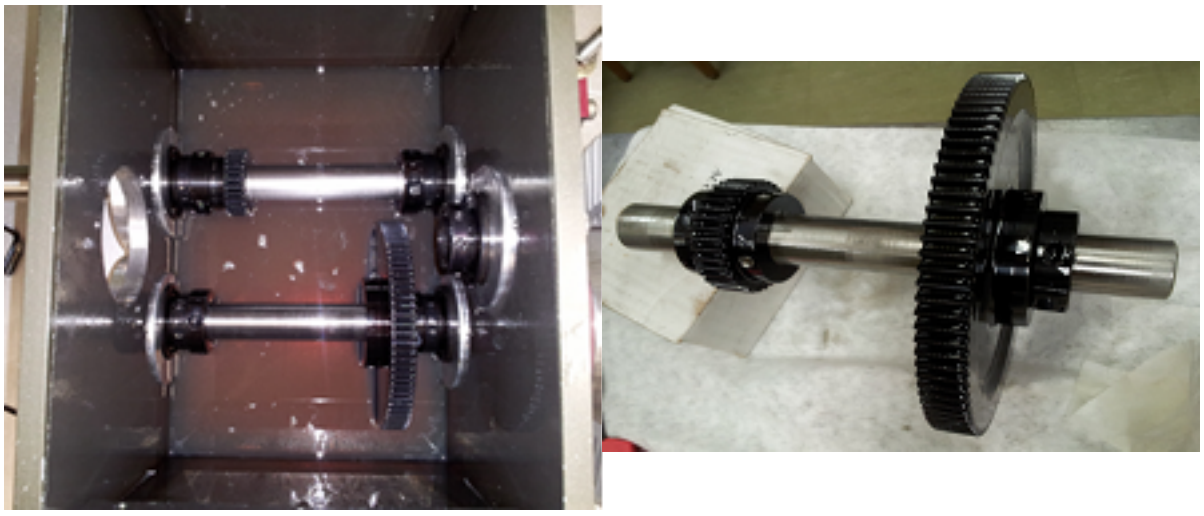
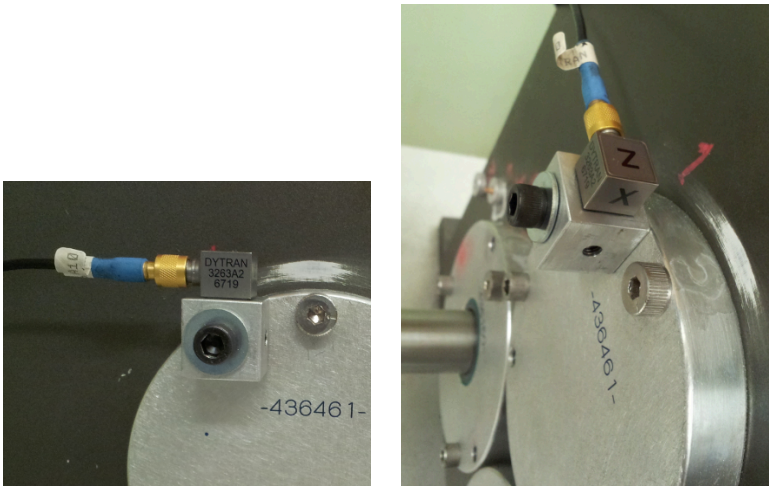


Figure 4.12: Intermediate Shaft Removed to Replace the Gear



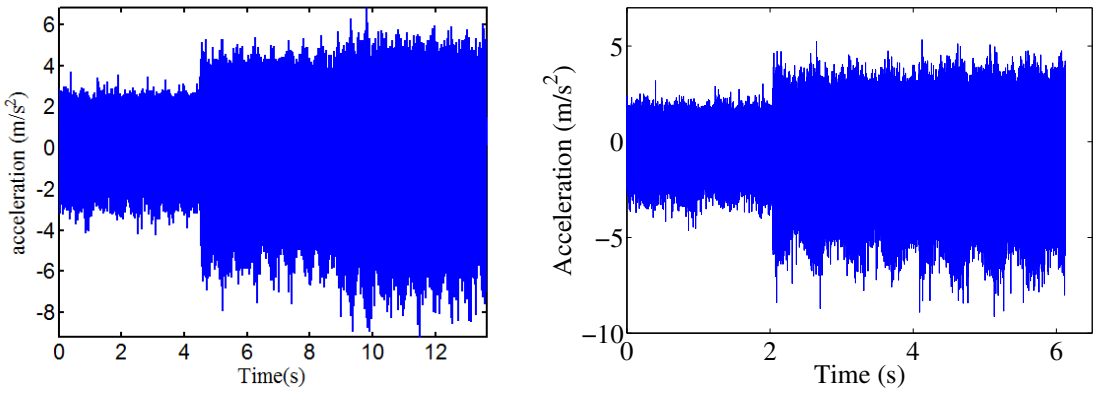
(a) Mounting Disk (b) Orientation

Figure 4.13: Accelerometer Mounted on a Mounting Disk

Figure 4.14: Lenze Controller

are carried out using the 36 teeth gear for the three cases of good, chipped, and missing tooth, which provides us with a 3 state data as a logical consequence. In the second set, tests are carried out under 3 degraded health conditions (4 overall health states): baseline data, followed by chipped tooth fault of the 36 teeth gear, chipped tooth fault with an outer race bearing fault, and finally chipped tooth fault with both inner and outer race faults on the bearings connected to the intermediate shaft. The datasets are concatenated in the aforementioned order to simulate a progressively deteriorating scenario. For all the test cases the operating condition is assumed to be steady (i.e. constant rpm). The sampling frequency used is 24 kHz.

4.2 Basic Signal Analysis



(a) Vibration Data - Sample 1

(b) Vibration Data - Sample 2

Figure 4.15: Acceleration Data for the 3 Health States

The case of three state data: good gear, chipped gear and missing tooth is considered first. Since triaxial accelerometers are used, three accelerations signals are recorded for

each test. However, only the X-direction (a radial direction, see Figure 4.13b) is found to be useful in terms of the information content (represented using CIs), and is hence retained for representation purposes. Figure 4.15 shows the 3 state acceleration data. It can be clearly observed from the Figure 4.15 that there is a significant difference in the amplitudes between the good and faulty state data; however there is no noticeable difference between the chipped tooth and the missing tooth cases. These differences become easier to differentiate when the spectral representation is investigated. Two samples of vibration data is presented here to show that a variation is observed between fault states for any number of samples of data collected. The vibration signal in Figure 4.15a is processed in the next section to interpret results.

The Fourier spectra of the acceleration vector is shown in Figure 4.16, corresponding to an input speed of 23 Hz. Multiple harmonics in the spectra can be clearly observed at the multiples of shaft speed at 24 Hz, for e.g., at 240 Hz and 667 Hz. GMFs corresponding to 480 Hz, 1680 Hz and 5281 Hz (i.e. multiples of 240 Hz) have significant energy compared to the multiple of 667 Hz where only the first two orders contain significant energy. The spectra of DDS with chipped gear (70% of the teeth dimension chipped) shows significant amount of sidebands and the appearance of harmonics at several frequencies other than the GMFs. This is attributable to the damage sustained owing to chipping. Similar nature can also be obtained for the spectra of the missing tooth gear, with the missing tooth case having widely separated sidebands and higher amplitude compared to the good and the chipped tooth counterparts.

Although the spectral differences between the good and the chipped tooth states (similarly missing tooth state as well) is obvious, this fact alone is not sufficient to automate the diagnostics. There are many extraneous peaks that can lead to errors in automation if spectral analysis is used as the sole discriminator. This issue is dealt with in the next

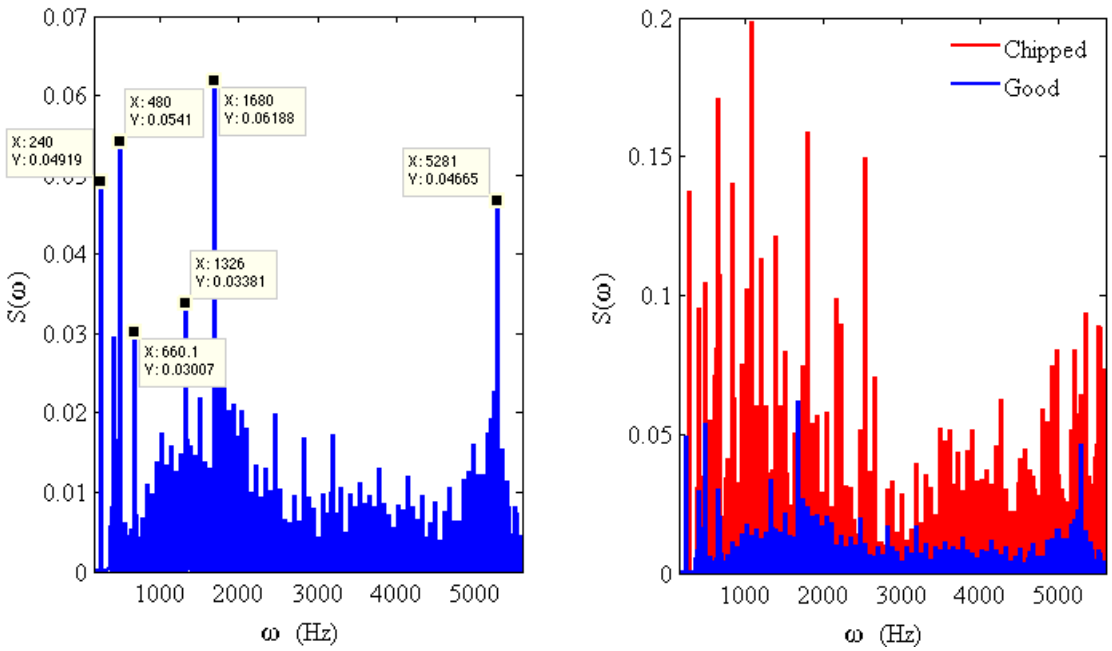


Figure 4.16: Fourier Spectra of the DDS Signal for Good and Chipped Tooth Conditions

Section.

4.3 Fault Diagnosis

4.3.1 Three State Data

Unlike synthetically generated data, experimental data is contaminated with electrical noise, which adds extraneous frequencies in the spectrum. Because of gear defects, shaft harmonics don't carry any useful information. They usually mask the gear meshing harmonics, which could sometimes have lower energy than the shaft harmonics. Hence, a pre-processing tool called empirical wavelet decomposition (EWD, see Appendix A for details) is used for de-noising. The case of input RPM 23Hz is considered first for illustration. This step essentially decomposes the signal into 20 single frequency components. Since the RPM is already known, the EWD components corresponding to the first 10 multiples of rpm are discarded along with some other components with high spectral energy that does not change with RPM. This is accomplished as window-wise data streams in. This is followed by the calculation of the features.

The signal $s(t)$ is of length 327000 and is sampled into windows of size 1000 each. The plots of each of $Y_{m\times l}$, where $l = 1, ..., n$, for m windows are shown in Figures 4.17, 4.18, 4.19 and 4.20. Standard deviation, RMS, amplitude square and root amplitude show that there is a possibility of 3 states in the data from the trend observed in their plots. Maximum and minimum shows the presence of 3 clusters, while kurtosis, normalized sixth moment, skewness, crest factor and pulse factor do not distinguish the three states clearly.

In the next step, dimensionality reduction using PCA is performed since working with 13 dimensional feature space becomes computationally expensive. The scatter plot between

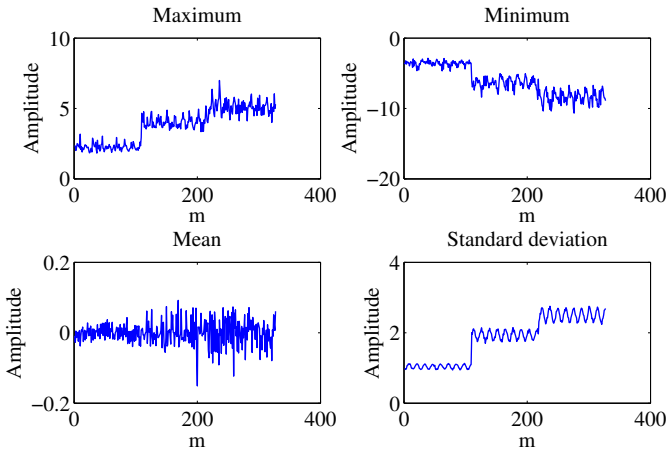


Figure 4.17: 3 States - Features Set 1

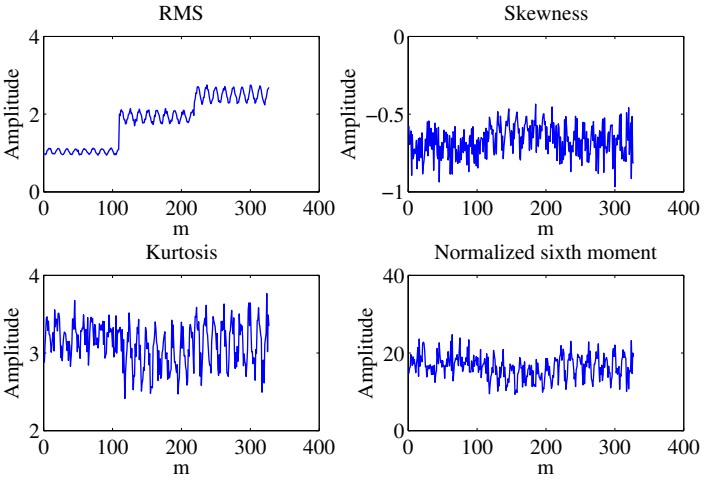


Figure 4.18: 3 States - Features Set 2

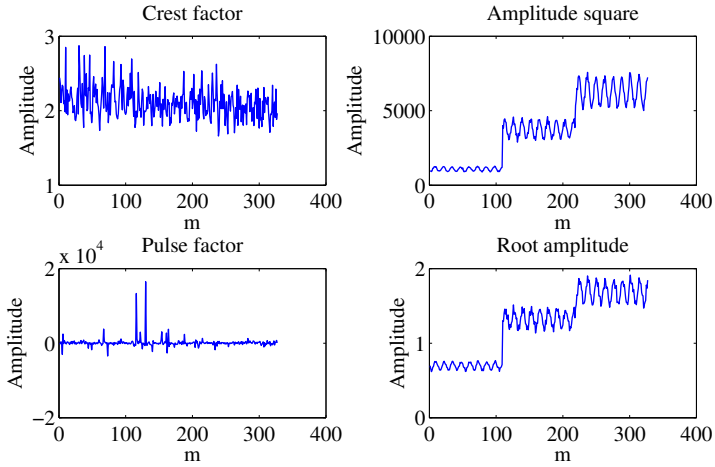


Figure 4.19: 3 States - Features Set 3

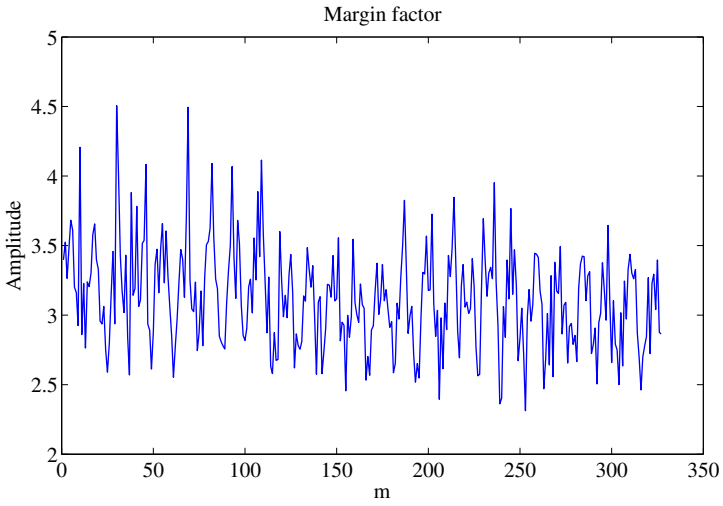


Figure 4.20: 3 States - Features Set 4

the principal components $PC1$ and $PC2$ is shown in Figure 4.21. The scatter plot shows that there are 3 separable health states.

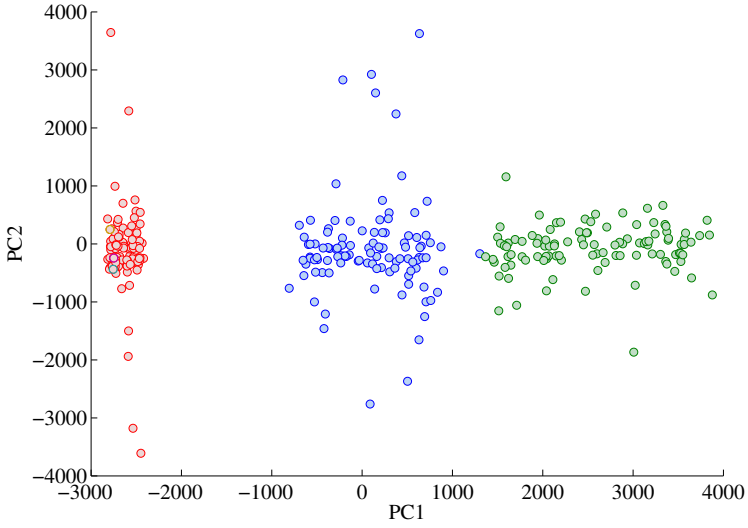


Figure 4.21: Scatter Plot

Figure 4.22 shows the centroids of the clusters obtained after k-means iteration. The number of clusters is further clarified through this procedure using the elbow phenomenon [57] as shown in Figure 4.23.

Once the number of clusters are determined, novelty detection, as explained in Section 3.4, is applied with logarithm of MD (η) as the novelty score.

Figure 4.24 shows the variation of η with index of data point. It can be observed from the Figure that near $l = 109$, there is a drastic change in the trend of η. So there are more than one fault conditions observed in the vibration signal, which is consistent with the test set-up. The fault trigger condition is set using SPC. For this purpose, the control rules are set using $\theta = 3$. It can be observed from the first control chart in Figure 4.25 that there is

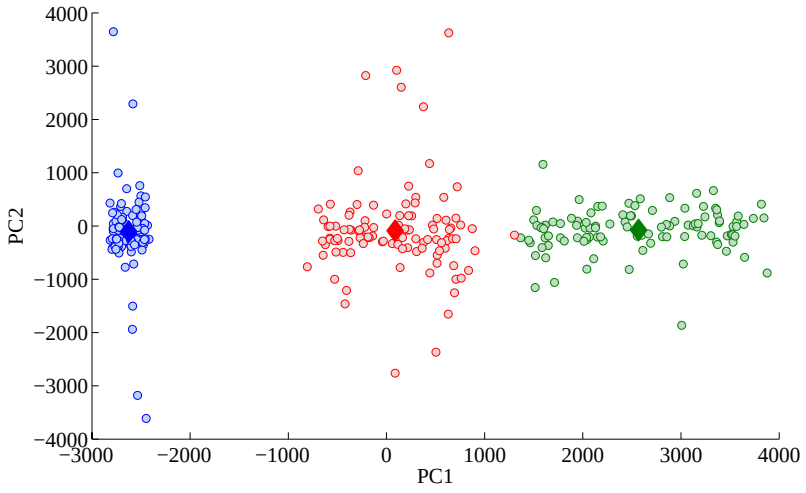


Figure 4.22: k-means Clustering

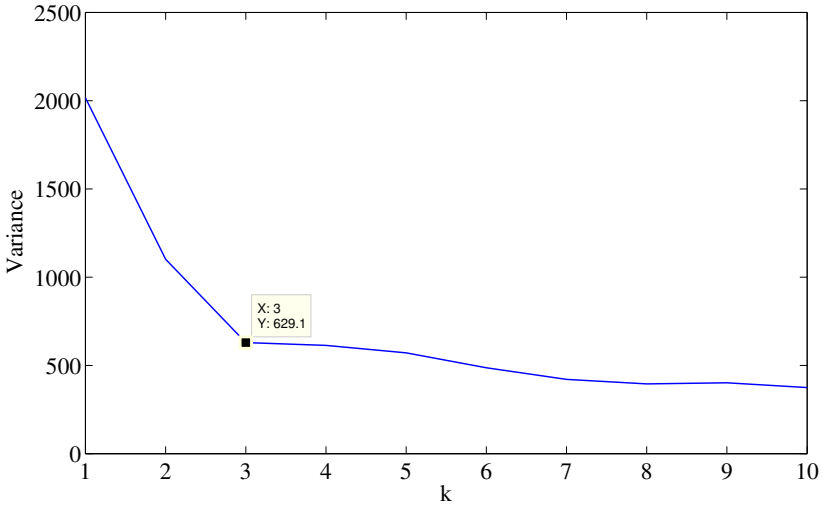


Figure 4.23: Number of Clusters

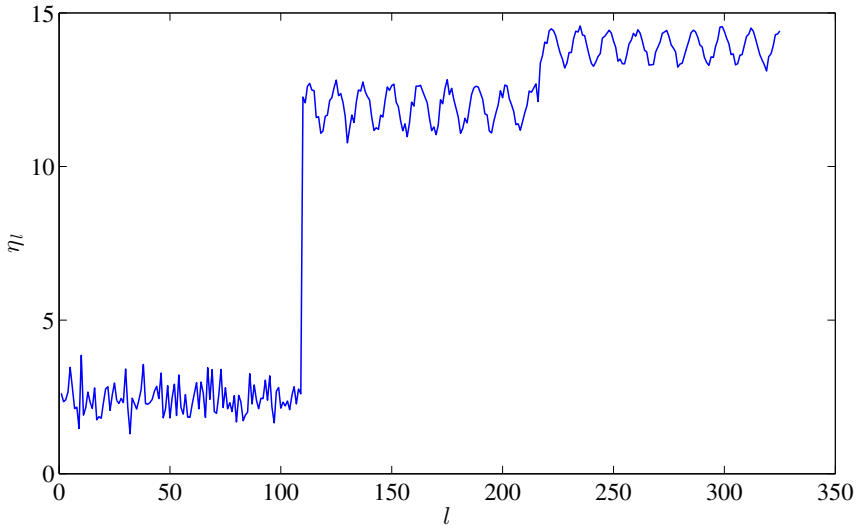


Figure 4.24: Novelty Score

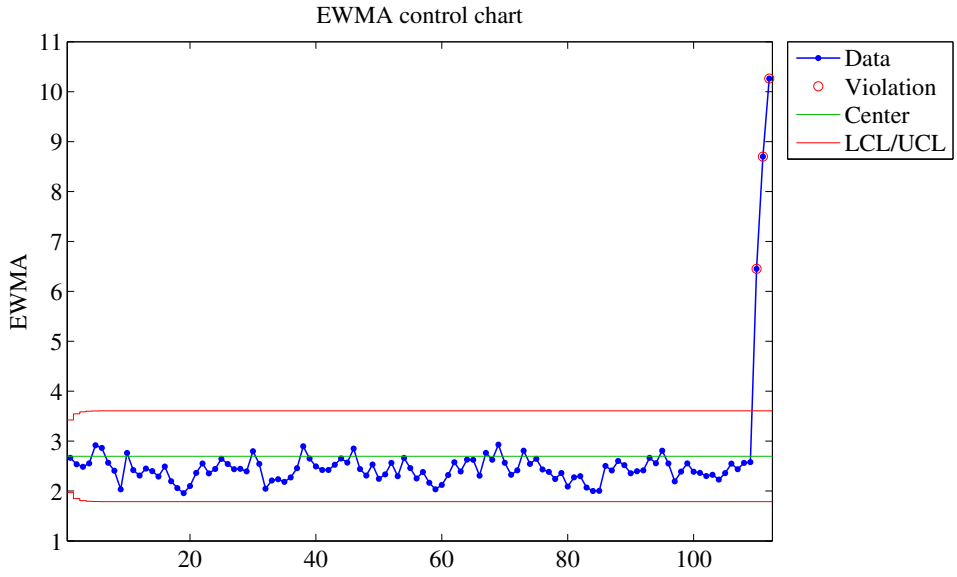


Figure 4.25: Monitoring Process - Until the First Alert

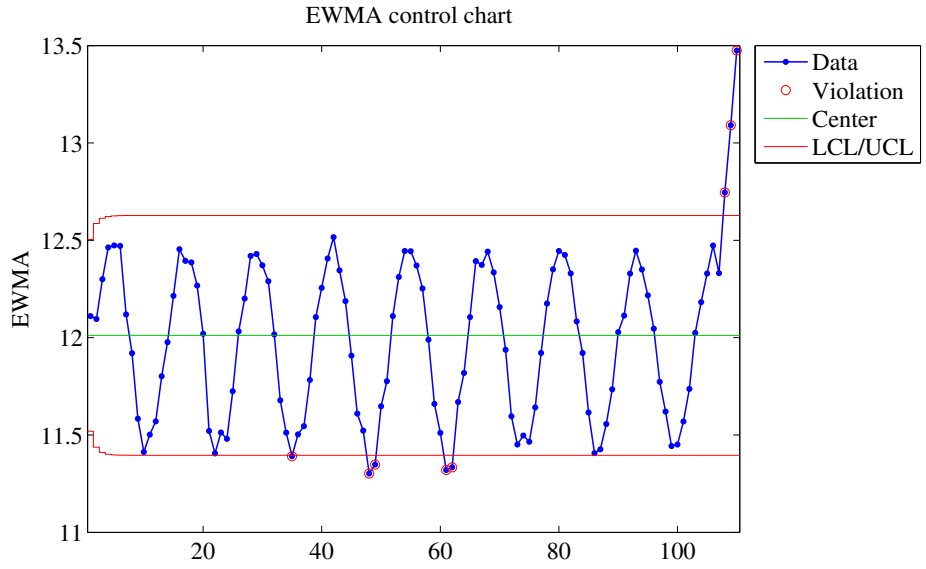


Figure 4.26: Monitoring Process - After the First Alert

a deviation in the process variable at $l = 109$. Therefore, at $l = 109$, the system triggers that the gearbox is behaving abnormally and it has to be monitored.

After the first trigger, a new data window is considered at $l = 109$ and the process variable and control limits are calculated again. At this stage, at $l = 218$, the system triggers a fault condition as shown in Figure 4.26. It is important to note that $l = 218$ corresponds to 21800 samples of data which is consistent with the introduction of the 3^{rd} health state of the gearbox (i.e. missing tooth condition). Thus, the framework of automated process control using clustering, novelty detection, and SPC successfully detects the 3 faulty states in the DDS data, consistent with the actual experimental conditions.

4.3.2 Four State Data

Now the case of faulty data with 4 states is considered. The only added complexity compared to the previous case arises from the introduction of faulty bearing in addition to gear faults. Figure 4.27 shows the 4 state acceleration data. It can be clearly observed from the Figure 4.27 that there is a significant difference in the amplitudes between the healthy state and the subsequent defective health states.

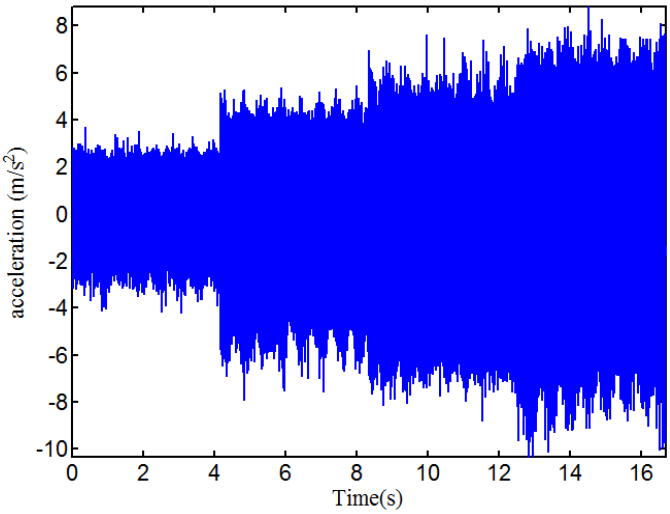


Figure 4.27: Acceleration Data for the 4 Health States

EWD is applied as a pre-processing de-noising tool to remove noise components corresponding to the first 10 multiples of RPM along with some other components with high energy spectral energy which do not change with RPM. This is accomplished window-wise as data is acquired. Then the features are calculated window-wise. The signal $s(t)$ is of length 392000. $s(t)$ is sampled into windows of size 1000 each. The plots of each of $Y_{m\times l}$, where $l = 1, ..., n$, for m windows are shown in Figures 4.28, 4.29, 4.30 and 4.31. Standard

deviation, RMS, amplitude square and root amplitude shows evidence of 4 states in the data from the trends, whereas the same cannot be concluded from the other condition indicators.

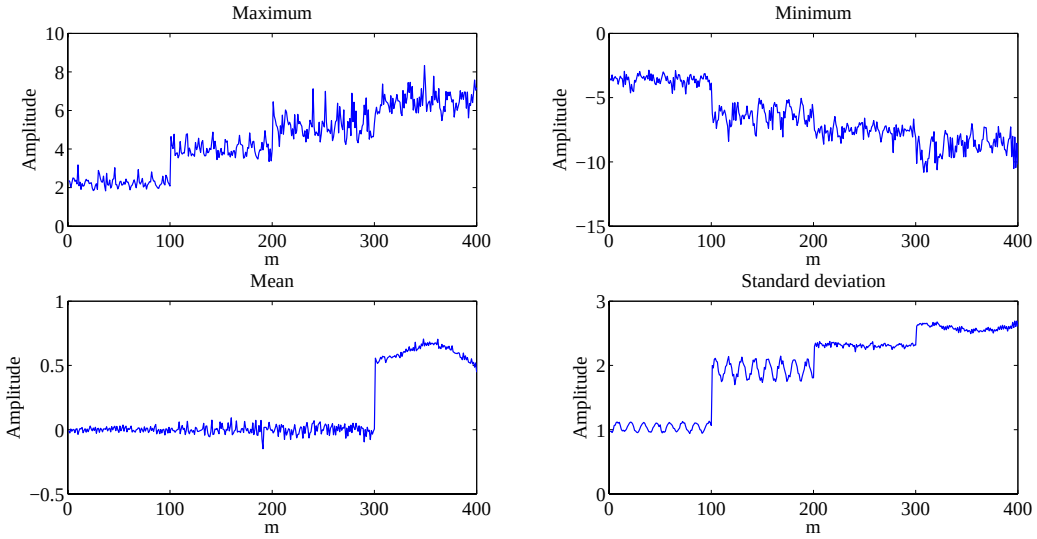


Figure 4.28: 4 States - Features Set 1

In the next step, dimensional reduction using PCA is performed. The scatter plot between the principal components $PC1$ and $PC2$ is shown in Figure 4.32. The scatter diagram shows that the data represents 4 separable health states.

Figure 4.32 also shows the centroids of the clusters obtained after k-means iteration. The number of clusters is further validated using the elbow principle as shown in Figure 4.33. Once the number of clusters is determined, novelty detection, as explained in Section 3.4, is applied to the data with logarithm of MD (η) as the novelty score.

Figure 4.34 shows the variation of η with index of data points, where η is calculated from the first centroid. It can be observed from the Figure that a drastic change is detected

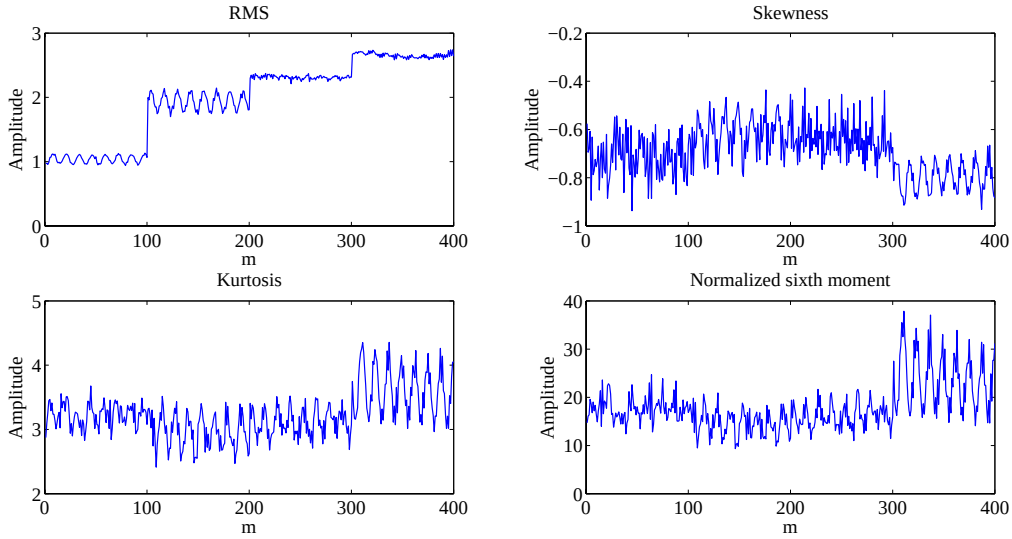


Figure 4.29: 4 States - Features Set 2

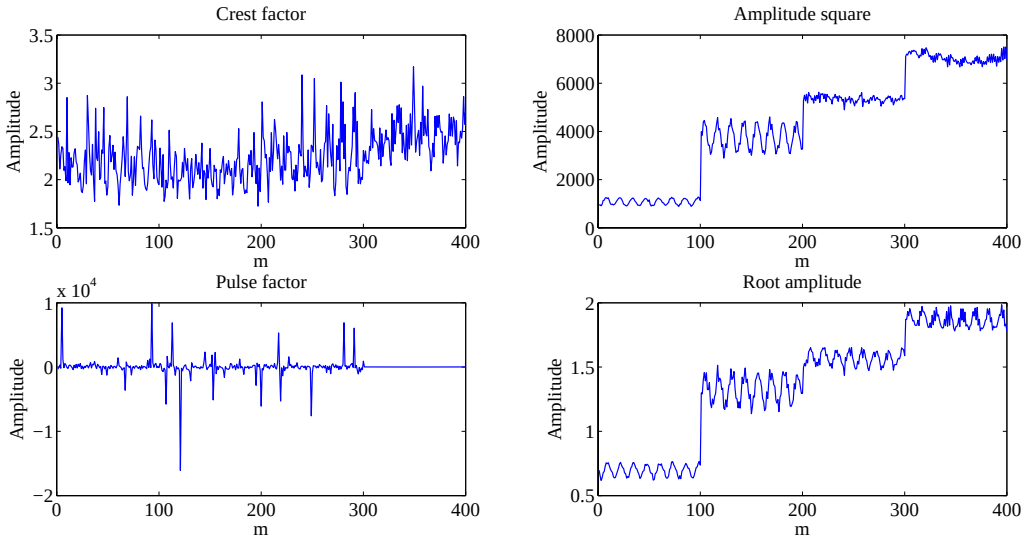


Figure 4.30: 4 States - Features Set 3

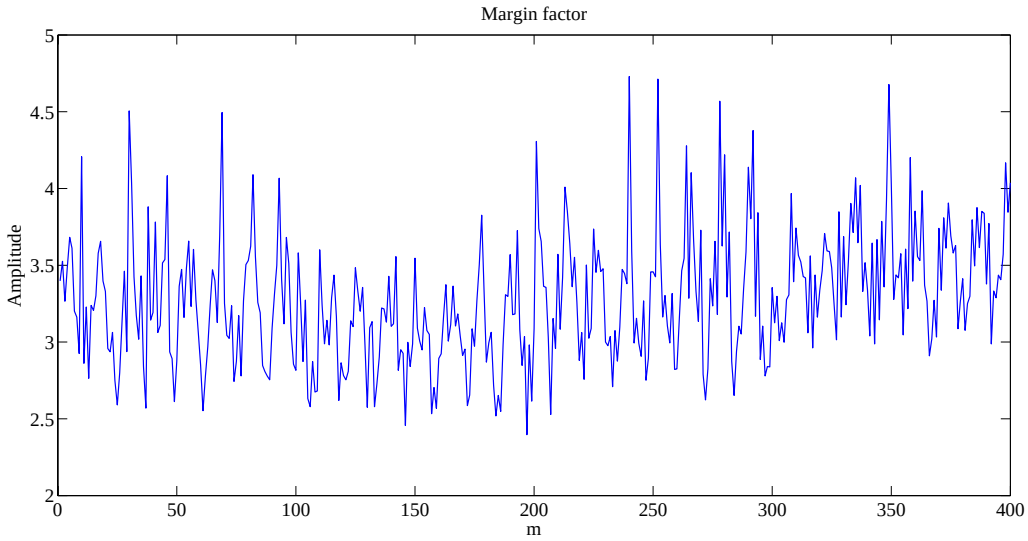


Figure 4.31: 4 States - Features Set 4

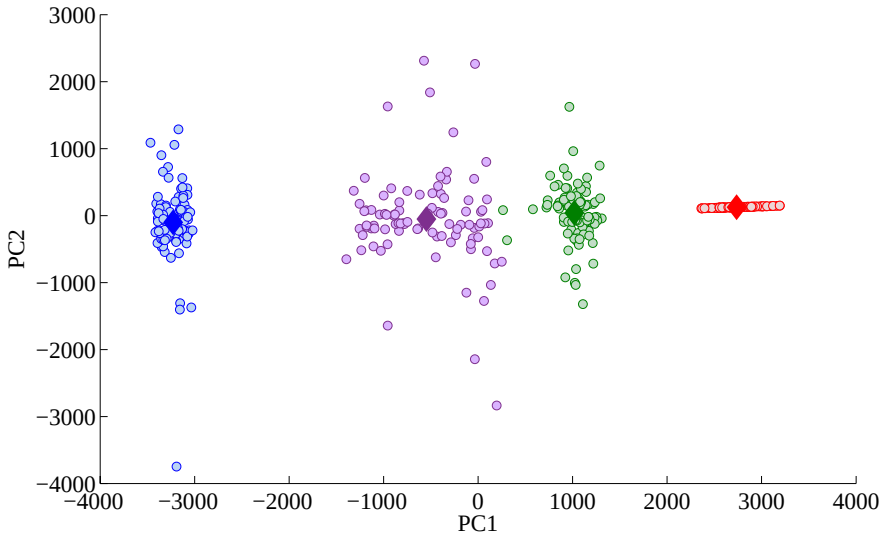


Figure 4.32: Clustering using k-means

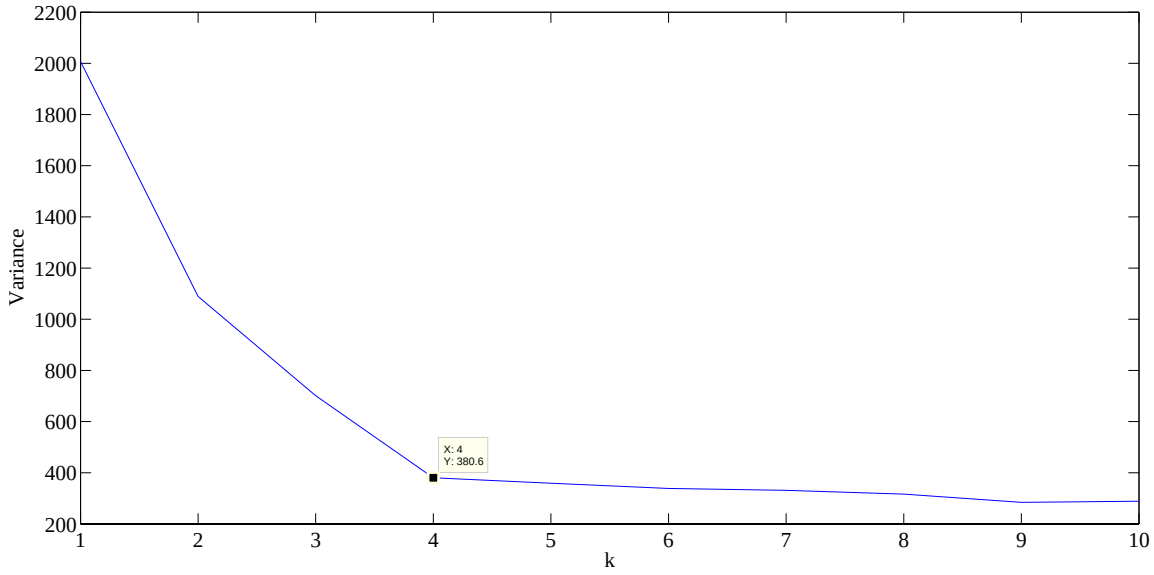


Figure 4.33: Number of Clusters

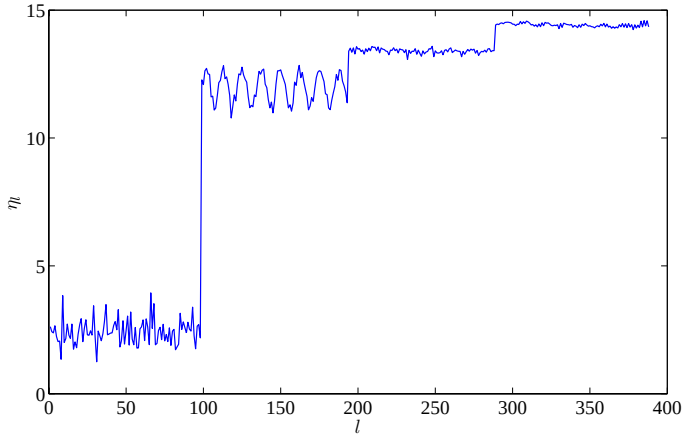


Figure 4.34: Novelty Score

in the trend of η near $l = 98$. So there are more than one fault conditions observed in the vibration signal, which is consistent with the experiment conducted. Proceeding to SPC, the control rules are set using $\theta = 3$. It can be observed from the first control chart in Figure 4.35 that there is a deviation in the process variable at $l = 98$. Therefore, at $l = 98$, the system triggers that the gearbox is behaving abnormally and the monitoring process starts.

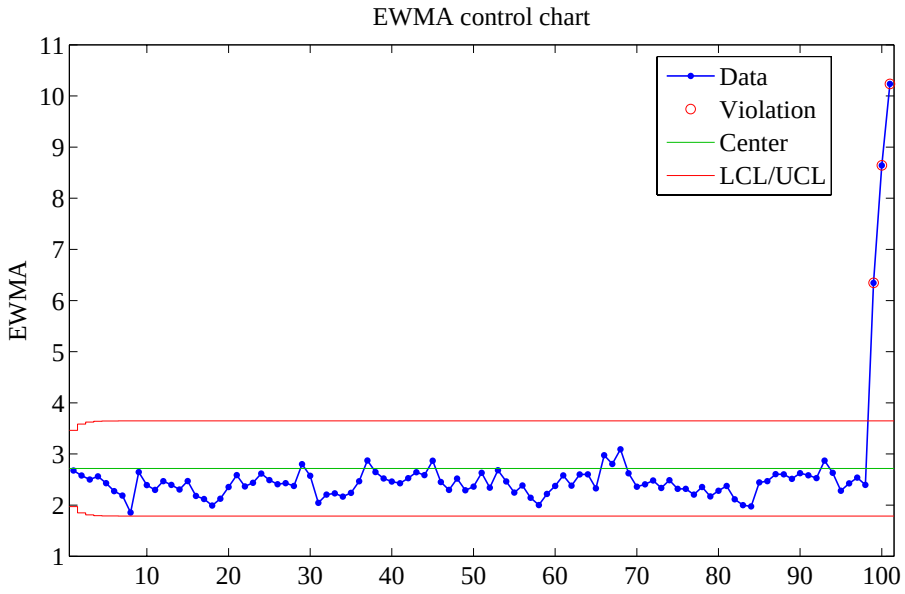


Figure 4.35: Monitoring Process - Until the First Alert

Subsequently, a new window is considered at $l = 98$ and the process variable and control limits are calculated. At this stage, at $l = 192$, the system triggers a fault condition as shown in Figure 4.36. In a similar manner, a third change is detected at $l = 288$. Thus, SPC is successfully able to detect 3 changes in the data that corroborate with the 4 health state data in the experiment Thus, it can be concluded that the framework of automated

process control using clustering, novelty detection and SPC successfully detects 4 faulty states in complex combination of gear and bearing faults which makes this algorithm an ideal candidate for gearbox condition monitoring in a practical setting.

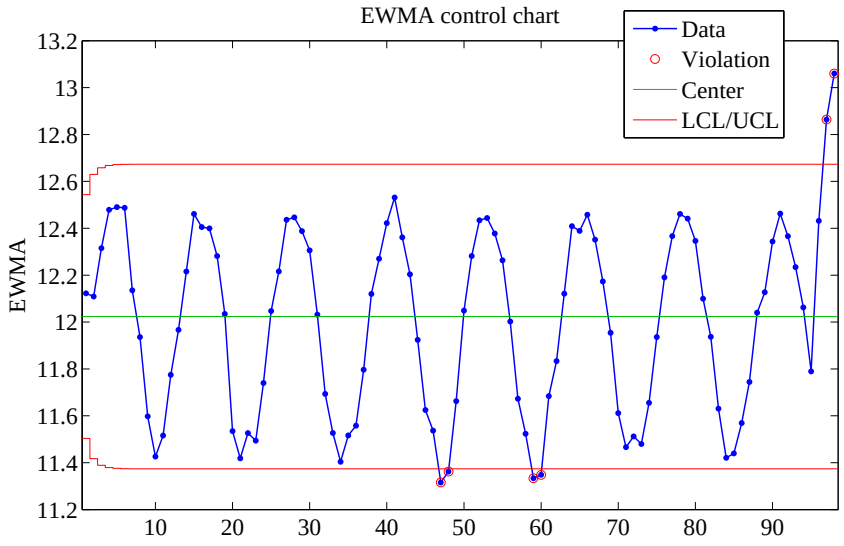


Figure 4.36: Monitoring Process - After the First Alert

Chapter 5

Summary, Conclusions and Future Work

5.1 Summary

A detailed study of automated condition monitoring techniques for gearbox diagnosis has been undertaken in this thesis, including review of traditional condition monitoring and statistical process control techniques, to develop a set of new techniques that can be automated for field applications. En-route, several targets have been accomplished in the context of automated gearbox condition monitoring. These are summarized as follows:

1. A automated gearbox diagnosis framework has been developed amalgamating basic ideas of condition monitoring, fault diagnosis, unsupervised learning, signal processing and statistical process control techniques.

2. The algorithm is capable of diagnosing faults online as data is collected or gathered

in real time where a change in the behaviour corresponds to a potential fault. Application of condition indicators, clustering, novelty detection and SPC all happens online with minimum manual intervention like comparing the output variables in traditional methods.

3. The algorithm is capable of detecting faults correctly for simulated as well as experimental vibration signals where the faults are mapped to instantaneous changes in condition. Results show robust detection of faults for experimental data obtained from the drivetrain dynamic simulator, both for a 3 health state case as well as a complex 4 health state case comprising a combination of gear and bearing faults.

4. The drivetrain dynamics simulator, a product of Spectraquest Inc, has been studied in detail. The details of operating the machinery and including stepwise gear and bearing replacement routines have been explained in detail to aid future research using this equipment for gearbox dynamics and condition monitoring.

5.2 Conclusions

Having highlighted the significant contributions of this dissertation, the central conclusions in this dissertation are summarized as follows:

1. Machine learning algorithms, in particular the unsupervised ones, are powerful tools for visualizing, representing and detecting faults in gearbox data, which are frequently represented by condition indictors. They are capable of automating the process of fault diagnosis significantly, which is an important step in fault diagnostics.

2. Statistical data driven approaches towards gearbox diagnostics have their shortcomings. They are heavily data dependent and their performance is only as good as the quality of the data. Signal pre-processing tools can improve the signal quality, thereby improving the performance of data driven algorithms.

3. Features have been extracted from the vibration signals and it is concluded that specific kind of features cannot be relied upon for fault diagnosis of gearbox. Using the available features, they need to transformed to identify a principal component space for analysis.

4. GMM clustering has been used to model pristine state data and this way, other health states are compared to the model to quantify the faulty condition.

5. Statistical process control is a novel way of automating the fault diagnosis of gearboxes because of its ability to generate alarms using online data. There are minimal adjustable parameters and hence, minimum manual intervention is required to accomplish fault diagnostics of gearboxes. It has been shown that this framework is sensitive to fault inception and helps in early prevention.

5.3 Recommendations for Future Study

Based on the results of this study, some recommendations and possible extensions of the current work are summarized as follows:

1. The complete methodology has been explored for the steady RPM case. A natural extension of this will incorporate an application and modification to varying speed data.

2. The algorithm could be extended to address progressively degrading gearboxes where the nature of degradation is continuous and transition from one fault to another is not obvious.

3. Since the current framework is completely data driven, other types of data such as that obtained from encoders, acoustic data or thermal data could be investigated.

4. Towards more advanced application of DDS for simulating gearbox faults, extensions could be made to cater to situations such as backlash, braking and radial loads applied to the bearings.

APPENDICES

Appendix A

Empirical Wavelet Decomposition

Empirical wavelet decomposition (EWD) is a signal decomposition method that has recently emerged in the context of non-stationary signal processing. Based on the concept of wavelet decomposition, its basic objective is to provide a time frequency representation of a non-stationary signal more adaptively than discrete wavelet and wavelet packet transform which is constrained by a fixed decomposition ratio. Motivated by the idea of constructing wavelet basis functions in Fourier domain and combining it with automatic peak detection, it can also be used to extract the individual components of a non-stationary multi-component signal, akin to empirical mode decomposition (EMD). The rich mathematical structure based on the concept of filter-bank algorithms makes AWD one of the potential candidates for gearbox diagnosis, as such signals are frequently constituted out of multiple amplitude-modulated / frequency-modulated signals (AM-FM) and periodic impulses due to bearing ring impacts embedded in noise [26].

The basic idea of EWD follows exactly similar lines as wavelet decomposition in the framework of multi-resolution analysis. EWD can be defined as an inner product of a

signal with an orthogonal basis function which is also called as the mother wavelet. The wavelet transform of a signal $x(t)$ is a linear transform, defined as :

$$w_k^j(x) = \frac{1}{\sqrt{j}} \int_{-\infty}^{\infty} x(t)\, \psi_k^{j*}(t)\, dt \tag{A.1}$$

where the function ψ is commonly known as the mother wavelet and $*$ stands for complex conjugation. j and k denotes scale and translation parameters respectively. Thus, wavelet transform decomposes a signal $x(t)$ via basis functions, that are simply scaled and translated versions of the mother wavelet. The key point of difference is the development of wavelet basis functions in the frequency domain also called Mayer wavelets. The main motivation behind this approach is that the traditional discrete wavelet decomposition are based on successive application of subsampling and decimation with a fixed ratio of decomposition (2^-j where j is the scale level). In EWD, the basis are constructed in frequency domain by finding the peaks in Fourier domain (i.e. Fourier transform) using an automatic peak detection algorithms [26]. Thus it does not have a fixed decomposition ratio and on the contrary it is based on adaptively building filter banks based on the spectral peaks of a signal in Fourier domain. Based on the location of the peaks ω_n and ω_{n+1} the filterbanks can be defined in the region $[\omega_n, \omega_{n+1}]$ as:

$$\psi_n(\omega) = \begin{cases} 1 \qquad if\ \omega_n + \tau_n \leq |\omega| \ \leq \omega_{n+1} - \tau_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}\left(|\omega| - \omega_{n+1} + \tau_{n+1}\right)\right)\right] if\ \omega_{n+1} - \tau_{n+1} \leq |\omega| \ \leq \omega_{n+1} + \tau_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}\left(|\omega| - \omega_n + \tau_n\right)\right)\right] if\ \omega_n - \tau_n \leq |\omega| \ \leq \omega_n + \tau_n \\ 0 \end{cases} \tag{A.2}$$

where τ_n is the width of the filterbank around the frequency ω_n.

Based on the above definitions the decomposition and reconstruction relations can be

written as:

$$\begin{aligned} W_x^{\varepsilon}(n,t) &= \langle x, \psi_n \rangle = \int x(\tau)\overline{\psi_n(\tau - t)}d\tau \\ &= (X(\omega)\psi_n(\omega)) \\ x(t) &= \sum_{n=1}^{N} W_x^{\varepsilon}(n,t) * \psi_n(t) \end{aligned} \tag{A.3}$$

Appendix B

Self-organizing Maps

Self-organizing map (SOM) [35], is a type of artificial neural network which uses unsupervised learning to train models. It is widely applied in the visualization of nonlinear relations of multidimensional data. Some of its applications include rotating machinery diagnostics [58]. SOM is a two-dimensional map containing neurons at the grid points. Each neuron is represented by a prototype vector (also called model or codebook vector), having same size as the input data set. During training and visualization phase, each input vector is assigned to the most similar prototype vector, also called best-matching unit (BMU). The algorithm trains itself in such that input vectors with similar features are mapped to relatively closer BMUs. The BMUs are updated iteratively during the training steps by selecting the input vector randomly. A neighbourhood kernel, whose radius decreases with iterations, determines the influence of input vector on the neighbouring BMUs. Starting in rough learning phase has a big influence area and fast-changing BMUs and shifts gradually to a fine-tuning phase with small influence area and slowly adapting BMUs. This algorithm is referred to as a sequential training or basic SOM.

SOM has also been applied in novelty detection [37]. To illustrate the modification of the basic SOM for novelty detection, given training set X, containing N pristine state data points, SOM is trained to generate a set of BMUs, $w = \{w_k \mid k = 1, 2, ..., K\}, K \ll N$. The codebook vector $m(x)$ of an input vector x and the Voronoi region S_k of each codebook vector w_k are defined as follows,

$$m(x) = w_k \Leftrightarrow x \in S_k \tag{B.1}$$

if

$$\|w_k - x\|^2 < \|w_l - x\|^2, \forall l \neq k \tag{B.2}$$

Given a test pattern z, the Euclidean distance (quantization error) $e(z)$ between z and $m(z)$ is defined as:

$$e(z) = \|z - m(z)\|^2 \tag{B.3}$$

If this quantization error is greater than a threshold value, then the corresponding input vector is considered to be novel. To identify the threshold value, the quantization errors corresponding to the training vectors are computed and the 95^{th} percentile value of the quantization errors is set as a global threshold value. Quantization error from test pattern is computed and compared against the threshold to identify if the input vector is novel or normal.

Appendix C

Principal Component Analysis

Principal component analysis (PCA) [33] transforms a feature vector into a new coordinate system by a linear combination of the features. Given a feature vector $\nu_{1\times n}$, the transformed vector $u_{1\times n}$ is obtained using the transformation matrix $T_{n\times n}$.

$$u_{1\times n} = T'_{n\times n} \times \nu_{1\times n} \tag{C.1}$$

$u_{1\times n}$, the resulting transformed feature vector of i^{th} feature is in the new coordinate system defined by the transformation matrix $T_{n\times n}$ that contains linear combination of feature $\nu_{1\times n}$. For a series of features, the transformation equation is given by:

$$U = T' \times V \tag{C.2}$$

The transformation matrix is of dimension $n \times n$ from above notations. It consists of vectors representing linear combinations of the input feature vectors, represented as:

$$T = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n-1} \end{bmatrix} \tag{C.3}$$

Transformation matrix

It should be noted that in the transformation matrix, the c_i are in the direction of highest variance in the input matrix V and the variance is contained within as few transformed vectors as possible. The vectors in transformation matrix are orthogonal to each other and satisfy the equation:

$$c_i'.c_j = 0, \forall i \neq j \tag{C.4}$$

To calculate the transformation matrix T, the covariance of the feature vector is calculated as:

$$\Phi = \frac{1}{n-1}(\nu - \mu_\nu)(\nu' - \mu_\nu') \tag{C.5}$$

where μ_ν is the mean of the input feature vector ν. The eigenvectors of Φ represent axes in the new coordinate system and form the transformation matrix. The eigenvectors are sorted in the decreasing order of the eigenvalues, and hence the vector c_0 corresponds to the highest eigenvalue. This highest eigenvalue contributes the most variance to the overall variance. The features calculated using first few components contain most of the information and explain the process due to which the lower components can be ignored. The transformation matrix can be reduced to a dimension $n \times l(l \leq n)$ where l is the required number of components when using PCA to reduce dimensionality.

Appendix D

Mahalanobis Distance

Mahalanobis distance (MD) [44] is a measure of the distance between a data point $x_{n\times 1}$ and a distribution (D) of points. The idea of MD is multi-dimensional generalization of a measure of number of standard deviations of x from the mean of D. The MD is zero when x is at the mean of D, and increases with x moving away from the mean. MD measures the number of standard deviations of x from the mean of D along each of the principle component axes (see Appendix C). When these axes are rescaled to unit variance, the MD corresponds to the standard Euclidean distance in the transformed space. Because of this reason, MD does not have units, is scale-invariant and accounts for the correlation present in the dataset.

The Mahalanobis distance of a data point $x_{n\times 1}$ from a cluster with mean, $\mu_{n\times 1}$ and covariance matrix $S_{n\times n}$ is given by equation:

$$MD(x) = \sqrt{(x-\mu)' S^{-1} (x-\mu)} \tag{D.1}$$

Mahalanobis distance (also known as generalized squared inter-point distance) is also defined as a dissimilarity measure between two vectors x and y, belonging to a same

distribution with a covariance matrix S, the equation for which is given by:

$$MD(x, y) = \sqrt{(x - y)' S^{-1} (x - y)} \tag{D.2}$$

Appendix E

Gaussian Mixture Models

A multivariate Gaussian distribution is a generalization of the single variable normal distribution and is given by:

$$N(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\} \tag{E.1}$$

where x is a D-dimensional random variable, μ is the mean vector and Σ is the covariance matrix.

When observations in a data set belong to multiple distributions, the dataset is modelled as a mixture of probability densities [42]. The mixture models combine many probability distributions and describe a more complex probability distributions. The Gaussian mixture models are represented by:

$$p(x) = \sum_{k=1}^{K}\pi_k N(x|\mu_k,\Sigma_k) \tag{E.2}$$

Equation E.2 represents a linear mixture of Gaussian densities $N(x|\mu,\Sigma)$. The param-

eters π_k are called the mixing coefficients and must satisfy the criterion:

$$\sum_{k=1}^{K} \pi_k = 1 \tag{E.3}$$

and

$$0 \geq \pi_k \geq 1 \tag{E.4}$$

given that $N(x|\mu_k, \Sigma_k) \geq 0$ and $p(x) \geq 0$. When GMM is used for unsupervised learning, the dataset is fit to GMM by finding the parameters for the mixtures using expectation maximization (EM) algorithm (see Appendix F).

Appendix F

Expectation Maximization

Given a dataset and the number of clusters k, the GMM parameters, μ and Σ of the clusters can be found to describe the dataset. Expectation maximization (EM) algorithm is an efficient iterative refinement approach for this purpose [71]. It gives the maximum likelihood estimates of the probability distributions and fits the dataset to the GMM models. The likelihood is given by the expression:

$$L = \prod_{n} p(x_n) \tag{F.1}$$

The GMM model can be written as (see Appendix E):

$$p(x) = \sum_{k=1}^{K} p(k)N(x|\mu_k, \Sigma_k) \tag{F.2}$$

It has to be noted that the π_k is referred to as $p(k)$ in the above equation and hereafter to feel the intuition of probability more than a mixing coefficient in this Appendix. GMM assigns each observation a set of weights corresponding to its belongingness to each cluster, called the membership score, represented by the conditional probability $p(k|n)$, where n is

the datapoint, and k that the membership score corresponds to. The matrix containing $p(k|n)$ s called the responsibility matrix, the expression for which is given by:

$$p_{nk} = p(k|n) = \frac{p(x_n|k)p(k)}{p(x_n)} = \frac{N(x_n|\mu_k, \Sigma_k)p(k)}{p(x_n)} \tag{F.3}$$

EM algorithm is alternates between two steps:

Expectation (E) step:

Given that the parameters μ_k, Σ_k and $p(k)$ are known (these parameters are usually obtained by using k-means clustering, so that the parameters are initialized with reasonable likelihood making the EM algorithm converge quickly), the likelihood L and the membership scores $p(k|n)$ are calculated.

Maximization (M) step:

In the maximization step, the parameters μ_k, Σ_k and the mixing coefficients $p(k)$ are calculated using the expressions in equation given by:

$$\begin{aligned} \hat{\mu}_k &= \frac{\sum_n p(k|n)x_n}{\sum_n p_{n|k}} \\ \hat{\Sigma}_k &= \frac{\sum_n (x_n - \hat{\mu}_k)(x_n - \hat{\mu}_k)}{\sum_n p(k|n)} \\ \hat{p}(k) &= \frac{1}{N}\sum_n p(k|n) \end{aligned} \tag{F.4}$$

The parameters are estimated such that the likelihood L is maximized for each observation.

Appendix G

k-means Clustering

k-means clustering [27] is a method that partitions m data points into k clusters where each data point belongs to the cluster with the nearest mean (a prototype of the cluster). The clustering problem is computationally difficult (NP-hard); to address this, algorithms that use iterative refinement approach are employed to make them converge quickly to a local optimum. This approach is similar to the Gaussian mixture models (see Appendix E) that use expectation maximization (EM) (see Appendix F) algorithm to find the local optimum. k-means is a hard clustering problem unlike GMM, which is soft clustering. Given a set of data points $(x_1, x_2, \cdots, x_n)$, where each data point is a d-dimensional vector, k-means clustering partitions the n data points into k sets $S = \{S_1, S_2, \cdots, S_k\}, (k \leq n)$ such that the within-cluster sum of squares ($WCSS$), given by Equation G.1, is minimized.

$$WCSS = \underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \tag{G.1}$$

where μ_i is the mean of points in S_i.

A standard algorithm of k-means is the most common among the ones using an iterative

refinement technique. It is also referred to as Lloyd's algorithm. Given an initial set of k means $m_1, \cdots, m_k$ (among various initialization procedures available, randomly selecting k points from the given set is used to explain here), the algorithm runs by alternating between two steps:

Assignment step: Each data point is assigned to a cluster whose mean gives the least within-cluster sum of squares ($WCSS$). Since the sum of squares quantity is the squared Euclidean distance, it is the nearest mean to the data point. The step is summarized in the equation:

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \leq \left\| x_p - m_j^{(t)} \right\|^2 \forall j, 1 \leq j \leq k \right\} \tag{G.2}$$

where each data point, x_p is assigned to only one cluster, $S^{(t)}$ using the above criterion.

Update step: New means, m_i are calculated which are the centroids of the data points in the new clusters.

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S_i^{(t)}} x_j \tag{G.3}$$

As the calculation of mean uses least-squares estimation, this step minimizes the within-cluster sum of squares ($WCSS$) objective in the process. The algorithm is said to have converged when the cluster assignments do not change any further. Both steps seek to optimize the $WCSS$ objective, and since these partitions are finite, the algorithm must converge to a (local) optimum.

References

[1] François Auger and Patrick Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *Signal Processing, IEEE Transactions on*, 43(5):1068–1089, 1995. 13

[2] Behrad Bagheri, Hojat Ahmadi, and Reza Labbafi. Application of data mining and feature extraction on intelligent fault diagnosis by artificial neural network and k-nearest neighbor. In *Electrical Machines (ICEM), 2010 XIX International Conference on*, pages 1–7. IEEE, 2010. 20

[3] DC Baillie and J Mathew. A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 10(1):1–17, 1996. 13, 14

[4] W Bartelmus and R Zimroz. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing*, 23(5):1528–1534, 2009. 17

[5] Anna Bartkowiak and Radoslaw Zimroz. Data dimension reduction and visualization with application to multi-dimensional gearbox diagnostics data: comparison of several methods. *Solid State Phenomena*, 180:177–184, 2012. 17

[6] Erik Leandro Bonaldi, Levy Ely de Lacerda de Oliveira, Jonas Guedes Borges da Silva, Germano Lambert-Torresm, and Luiz Eduardo Borges da Silva. *Induction Motors - Modelling and Control*. InTech, 2012. xi, 3

[7] Great Britain and Michael Neale. *A guide to the condition monitoring of machinery*. HM Stationery Office, 1979. 8

[8] Tom Brotherton, Tom Pollard, and D Jones. Applications of time-frequency and time-scale representations to fault detection and classification. In *Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium*, pages 95–98. IEEE, 1992. 19

[9] Burgemeister. Accelerometer mounting, 1999. 11

[10] V Capdevielle, Ch Serviere, and JL Lacoume. Blind separation of wide-band sources: Application to rotating machine signals. In *Proc. of the 8th European Signal Processing Conf*, volume 3, pages 2085–2088, 1996. 14

[11] Leo H Chiang, Richard D Braatz, and Evan L Russell. *Fault detection and diagnosis in industrial systems*. Springer, 2001. 3

[12] Leon Cohen. *Time-frequency analysis*, volume 778. Prentice Hall PTR Englewood Cliffs, NJ:, 1995. 13

[13] Ronald R Coifman and M Victor Wickerhauser. Entropy-based algorithms for best basis selection. *Information Theory, IEEE Transactions on*, 38(2):713–718, 1992. 13

[14] EJ Cross, G Manson, K Worden, and SG Pierce. Features for damage detection with insensitivity to environmental and operational variations. *Proceedings of the*

Royal Society A: Mathematical, Physical and Engineering Science, 468(2148):4098–4122, 2012. 17

[15] Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Applied and computational harmonic analysis*, 30(2):243–261, 2011. 14

[16] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000. 19

[17] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977. 36

[18] JP Dron, L Rasolofondraibe, F Bolaers, and A Pavan. High-resolution methods in vibratory analysis: application to ball bearing monitoring and production machine. *International journal of solids and structures*, 38(24):4293–4313, 2001. 13

[19] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012. 17, 18, 19

[20] Ronald L Eshleman and Judith Nagle-Eshleman. *Basic Machinery Vibrations: An Introduction to Machine Testing, Analysis, and Monitoring*. VIPress, 1999. 10

[21] Xianfeng Fan and Ming J Zuo. Gearbox fault detection using hilbert and wavelet packet transform. *Mechanical Systems and Signal Processing*, 20(4):966–982, 2006. 2, 29

[22] Charles R Farrar and Keith Worden. *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, 2012. 16, 17, 18, 20, 22, 23, 42

[23] Dimitar P Filev, Ratna Babu Chinnam, Finn Tseng, and Pundarikaksha Baruah. An industrial strength novelty detection framework for autonomous equipment monitoring and diagnostics. *Industrial Informatics, IEEE Transactions on*, 6(4):767–779, 2010. 17, 23, 24

[24] DP Filev and F Tseng. Novelty detection based machine health prognostics. In *Evolving Fuzzy Systems, 2006 International Symposium on*, pages 193–199. IEEE, 2006. 22

[25] G Gelle, M Colas, and G Delaunay. Blind sources separation applied to rotating machines monitoring by acoustical and vibrations analysis. *Mechanical Systems and Signal Processing*, 14(3):427–442, 2000. 14

[26] Jérôme Gilles. Empirical wavelet transform. *Signal Processing, IEEE Transactions on*, 61(16):3999–4010, 2013. 5, 14, 82, 83

[27] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979. 95

[28] Budhaditya Hazra, Shilpa Pantula, and Sriram Narasimhan. Novelty detection in airport baggage conveyor gear-motors using synchro-squeezing transform and self-organizing maps. *PHM Society Conference*, 4(060), 2013. 22

[29] Qingbo He, Fanrang Kong, and Ruqiang Yan. Subspace-based gearbox condition monitoring by kernel principal component analysis. *Mechanical Systems and Signal Processing*, 21(4):1755–1772, 2007. 18

[30] Qingbo He, Ruqiang Yan, Fanrang Kong, and Ruxu Du. Machine condition monitoring using principal component representations. *Mechanical Systems and Signal Processing*, 23(2):446–466, 2009. 18

[31] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998. 14

[32] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006. 8, 12

[33] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005. 87

[34] Manabu Kano, Shouhei Tanaka, Shinji Hasebe, Iori Hashimoto, and Hiromu Ohno. Monitoring independent components for fault detection. *AIChE Journal*, 49(4):969–976, 2003. 23

[35] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990. 85

[36] Mitchell Lebold, Katherine McClintic, Robert Campbell, Carl Byington, and Kenneth Maynard. Review of vibration analysis methods for gearbox diagnostics and prognostics. In *Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology*, volume 634, 2000. 16

[37] Hyoung-joo Lee and Sungzoon Cho. Som-based novelty detection using novel data. In *Intelligent Data Engineering and Automated Learning-IDEAL 2005*, pages 359–366. Springer, 2005. 86

[38] Joon-Hyun Lee, J Kim, and Han-Jun Kim. Development of enhanced wigner–ville distribution function. *Mechanical systems and signal processing*, 15(2):367–398, 2001. 13

[39] Sun Ung Lee, David Robb, and Colin Besant. The directional choi–williams distribution for the analysis of rotor-vibration signals. *Mechanical Systems and Signal Processing*, 15(4):789–811, 2001. 13

[40] Younjeong Lee, Ki Yong Lee, and Joohun Lee. The estimating optimal number of gaussian mixtures based on incremental k-means for speaker identification. *International Journal of Information Technology*, 12(7):13–21, 2006. 36

[41] Weihua Li, Tielin Shi, Guanglan Liao, and Shuzi Yang. Feature extraction and classification of gear faults using principal component analysis. *Journal of Quality in Maintenance Engineering*, 9(2):132–143, 2003. 18, 19

[42] Bruce G. Lindsay. Mixture models: Theory, geometry and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 5:pp. i–iii+v–ix+1–163, 1995. 91

[43] Huageng Luo, Hai Qiu, George Ghanime, Melinda Hirz, and Geo van der Merwe. Synthesized synchronous sampling technique for differential bearing damage detection. *Journal of Engineering for Gas Turbines and Power*, 132(7):072501, 2010. xi, 4

[44] Prasanta Chandra Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936. 19, 89

[45] Arnaz Malhi and Robert X Gao. Pca-based feature selection scheme for machine defect classification. *Instrumentation and Measurement, IEEE Transactions on*, 53(6):1517–1525, 2004. 18

[46] Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007. 23, 42

[47] NG Nikolaou and IA Antoniadis. Rolling element bearing fault diagnosis using wavelet packets. *Ndt & E International*, 35(3):197–205, 2002. 13

[48] Linilson R Padovese. Using acoustical noise for fault classification in gearbox. In *Proceedings of the 15th Brazilian Congress of Mechanical Engineering, Sao Paulo, Brazil*, 1999. 19

[49] A Parey, Mohamed El Badaoui, François Guillet, and N Tandon. Dynamic modelling of spur gear pair and application of empirical mode decomposition-based statistical analysis for early detection of localized tooth defect. *Journal of sound and vibration*, 294(3):547–561, 2006. 14

[50] B Eugene Parker Jr, Todd M Nigro, Monica P Carley, Roger L Barron, David G Ward, H Vincent Poor, Dennis Rock, and Thomas A DuBois. Helicopter gearbox diagnostics and prognostics using vibration signature analysis. In *Optical Engineering and Photonics in Aerospace Sensing*, pages 531–542. International Society for Optics and Photonics, 1993. 19

[51] Robert Bond Randall. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011. 7, 9, 10, 11, 29

[52] BKN Rao. Handbook of condition monitoring, 1996. 8

[53] Cornelius Scheffer and Paresh Girdhar. *Practical machinery vibration analysis and predictive maintenance*. Elsevier, 2004. 2

[54] WJ Staszewski. Wavelet based compression and feature selection for vibration analysis. *Journal of sound and vibration*, 211(5):735–760, 1998. 13

[55] WJ Staszewski, K Worden, and GR Tomlinson. Time–frequency analysis in gearbox fault detection using the wigner–ville distribution and pattern recognition. *Mechanical systems and signal processing*, 11(5):673–692, 1997. 13

[56] RM Stewart. *Some useful data analysis techniques for gearbox diagnostics*. University of Southampton, 1977. 17

[57] Robert L Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953. 36, 66

[58] Markus Timusk, Mike Lipsett, and Chris K Mechefske. Fault detection using transient machine signals. *Mechanical Systems and Signal Processing*, 22(7):1724–1749, 2008. 2, 18, 21, 85

[59] Tim Toutountzakis and David Mba. Observations of acoustic emission activity during gear defect diagnosis. *NDT & E International*, 36(7):471–477, 2003. 17

[60] John Joseph Uicker, Gordon R Pennock, Joseph Edward Shigley, and J Michael Mccarthy. *Theory of machines and mechanisms*. Oxford University Press New York, 2003. 7

[61] George Vachtsevanos, Frank Lewis, Michael Roemer, Andrew Hess, and Biqing Wu. Intelligent fault diagnosis and prognosis for engineering systems. *Usa 454p Isbn*, 13:978–0, 2006. 2

[62] P Večeř, Marcel Kreidl, and Radislav Šmíd. Condition indicators for gearbox condition monitoring systems. *Acta Polytechnica*, 45(6), 2005. 3, 17

[63] Wenyi Wang and Albert K Wong. Autoregressive model-based gear fault diagnosis. *Journal of Vibration and Acoustics*, 124(2):172–179, 2002. 13

[64] WJ Wang and PD McFadden. Application of orthogonal wavelets to early gear damage detection. *Mechanical Systems and Signal Processing*, 9(5):497–507, 1995. 13

[65] John H Williams, Alan Davies, and Paul R Drake. *Condition-based maintenance and machine diagnostics*. Springer, 1994. 8

[66] MLD Wong, LB Jack, and AK Nandi. Modified self-organising map for automated novelty detection applied to vibration signal monitoring. *Mechanical Systems and Signal Processing*, 20(3):593–610, 2006. 22

[67] Keith Worden. Structural fault detection using a novelty measure. *Journal of Sound and vibration*, 201(1):85–101, 1997. 3

[68] Keith Worden and JM Dulieu-Barton. An overview of intelligent fault detection in systems and structures. *Structural Health Monitoring*, 3(1):85–98, 2004. 3, 18, 23

[69] Keith Worden and Graeme Manson. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):515–537, 2007. 19

[70] Keith Worden, Graeme Manson, and NRJ Fieller. Damage detection using outlier analysis. *Journal of Sound and Vibration*, 229(3):647–667, 2000. 3

[71] Lei Xu and Michael I Jordan. On convergence properties of the em algorithm for gaussian mixtures. *Neural computation*, 8(1):129–151, 1996. 93

[72] Alexander Ypma, Amir Leshem, and Robert PW Duin. Blind separation of rotating machine sources: bilinear forms and convolutive mixtures. *Neurocomputing*, 49(1):349–368, 2002. 14

[73] YM Zhan and AKS Jardine. Adaptive autoregressive modeling of non-stationary vibration signals under distinct gear states. part 1: modeling. *Journal of Sound and Vibration*, 286(3):429–450, 2005. 12, 13

[74] Radoslaw Zimroz and Anna Bartkowiak. Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing*, 38(1):237–247, 2013. 18