University of Strathclyde
Department of Electronic and Electrical Engineering

Developing Anomaly Detection, Diagnostics, and Prognostics for Condition Monitoring with Limited Historical Data in New Applications such as Tidal Power

by

Grant Stewart Galloway

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2017

Abstract

Tidal power is a promising source of renewable energy worldwide, more stable and predictable than alternatives such as offshore wind power. However, the harsh operating environment makes maintenance of tidal turbines difficult and costly to perform. Intelligent condition monitoring systems, utilised as part of a condition-based maintenance strategy, can provide operators with timely and accurate indications of faults before serious damage occurs. Nevertheless, tidal technology is a new application, where deployments are currently in its infancy. Therefore, there is limited practical experience of how faults will develop within tidal turbines in operation. This thesis aims to investigate the requirements of condition monitoring methods, from both theory and knowledge of similar fields (such as offshore wind), to develop an approach for applying condition monitoring in new applications. This work first investigates the response of a commercial-scale tidal turbine in operation through a data mining analysis of condition data from the Andritz Hydro Hammerfest HS1000 tidal turbine. Data mining was performed following the CRISP-DM methodology and was used to build models of the normal condition response of turbine components, from which anomalous behaviour indicative of the development of faults can be detected. This approach was then expanded to include both diagnostic and prognostic modelling, where faults can be automatically classified and the remaining useful life of equipment undergoing degradation can be estimated. This has resulted in a generalised framework, based on CRISP-DM, that can be applied to perform condition monitoring in new applications, learning from the response of machinery over time during its operation.

Contents

List of Figures

List of Tables

Abbreviations

ANN	Artificial Neural Network
BPFI	Ball Pass Frequency of Inner-raceway
BPFO	Ball Pass Frequency of Outer-raceway
BSF	Ball Spin Frequency
CI	Confidence Interval
CRA	Cumulate Relative Accuracy
CRISP-DM	Cross-Industry Standard Process for Data Mining
CUSUM	Cumulative Sum
DFIG	Doubly-fed Induction Generator
DTFT	Discrete-Time Fourier Transform
EKF	Extended Kalman Filter
EM	Expectation-Maximisation
EMEC	European Marine Energy Centre
FFNN	Feed Forward Neural Network
FFT	Fast Fourier Transform
FMEA	Failure Modes and Effects Analysis
FTF	Fundamental Train Frequency
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
KDE	Kernel Density Estimation
k-NN	k-Nearest Neighbours
PCA	Principal Component Analysis
PDF	Probability Density Function
PHM	Prognostics and Health Management Society
PLC	Programmable Logic Controller
RA	Relative Accuracy
RBPF	Rotor Bar Pass Frequency
RMS	Root Mean Square
RMSE	Root Mean Squared Error
RPM	Rotations per Minute
RUL	Remaining Useful Life
SCADA	Supervisory Control and Data Acquisition
SIR	Sampling Importance Resampling
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
UKF	Unscented Kalman Filter
WMEP	Wind Measurement and Evaluation Program

Nomenclature

Chapter 3 Anomaly Detection through Data Mining

x	Input data parameter
y	Output data parameter
cov	covariance
i	data point index
N	number of data points
r_{xy}	Pearson's correlation coefficient between parameters x and y
p	number of parameters in dataset
σ^2	variance
$\mathbf{V}$	covariance matrix
$\mathbf{X}$	input data matrix
$\mathbf{I}$	identity matrix
λ	Lagrange multiplier (principal component analysis)
C_i	Cumulative sum of differences at point i
μ	mean
$\mathbf{\Sigma}$	covariance matrix of Gaussian model
π	weight of model/parameter
$\boldsymbol{\theta}$	set of model parameters
$P(k\|x)$	probability of k given x
T	threshold value
$K(\cdot)$	kernel function
h	smoothing parameter (kernel density estimation)
$\boldsymbol{c}_k$	kth centroid (k means)
$\boldsymbol{Z}_p$	z score of pth dimension
$f(x,\theta)$	parametric Function
r	residual signal
S	sum of squared errors
ε	noise or error parameter
w	weight of

Chapter 4 Diagnostic Modelling

f_{mesh}	gear mesh frequency
f_{gear}	gear rotation frequency
f_{pinion}	pinion rotation frequency
f_{planet}	planet gear rotation frequency
f_{carrier}	carrier gear rotation frequency
$f_{\mathrm{planet-pass}}$	planet gear pass frequency
$f_{\mathrm{planet-mesh}}$	planet gear mesh frequency
f_{s}	planet gear rotation frequency
$N_{\mathrm{P,S,R}}$	Number of gear teeth in planet, sun, ring gear respectively
M_{P}	Number of planet gears
f_{R}	Rotation frequency

G	highest common factor between number of gear and pinion teeth
f_a	amplitude frequency
f_c	carrier frequency
$f_\text{hunting tooth}$	hunting tooth frequency
f_natural	natural resonance frequency
Bd	bearing ball diameter (mm)
Pd	bearing pitch diameter (mm)
θ_contact	bearing contact angle
Nb	number of bearing rolling elements
f_slip	slip frequency
f_p	pole-pass frequency
f_rotor	rotor frequency
f_L	line frequency
n_poles	Number of generator poles
n_bars	Number of rotor bars
N_C	Critical shaft speed
ϕ	order of shaft rotation
x_RMS	root mean square of signal x
x_peak	peak of signal x
C	crest factor
K	kurtosis
$x_\text{p-p}$	peak to peak amplitude of signal x
d	difference signal
E	envelope signal
$X(f)$	Fourier Transform of time domain signal $x(t)$
$\Psi(\cdot)$	Mother wavelet function
$W(a;b)$	Wavelet transform with scaling factor a and time-shift parameter b
$C(\tau)$	Cepstrum signal
$I_G(p)$	Gini impurity index
$d(\cdot)$	Distance function
a	Activation of neuron in neural network
W	Weight of neuron in neural network
b	Bias of neuron in neural network
K	Number of training examples
$\delta^{(L)}$	Error term at layer L in neural network
η	Gradient descent scaling factor
Ω_weights	Weight decay
Ω_sparsity	Sparsity constant
$\hat{\rho}$	Average activation of hidden neuron
$\text{KL}(\cdot)$	Kullback-Leibler divergence
Z'	Z-score standardisation

Chapter 5 Prognostics

$p(\cdot)$	Probability distribution
β	Weibull distribution shape parameter

λ	Weibull distribution scale parameter
k	Time point
k_{E}	Point of failure
T_{E}	Failure threshold
$\boldsymbol{\theta}$	Degradation model parameters
$\mathbf{F}$	State transition model
$\mathbf{B}$	Control-input model
$\boldsymbol{u}$	Operational parameters
$\boldsymbol{\omega}$	Process noise
$\mathbf{Q}$	Covariance of process noise
$\mathbf{H}$	Observation model
$\boldsymbol{v}$	Observation noise
$\mathbf{R}$	Covariance of observation noise
$\mathbf{P}$	State estimate error covariance
$\mathbf{K}$	Kalman gain
$\boldsymbol{\chi}$	Sigma points
$\boldsymbol{\gamma}$	Transformed sigma points
P	Number of particles
P_{eff}	Effective number of particles
ξ	Noise parameter
F_k	Fault signal at step k
v_k	Vibration signal at step k
ϵ	Gaussian noise

Chapter 1

1 Introduction

1.1 Introduction to Research

The aim of this thesis is to present an approach for applying condition monitoring in new applications, where there is limited historical monitoring data or knowledge of the system.

A key application of this work is tidal power, a newly developing form of renewable energy generation, capturing tidal forces to provide a reliable and predictable source of energy. This technology has great potential to become a major source of renewable energy, within the UK and worldwide [1].

Maintaining equipment within the marine environment is a major challenge. Firstly, access to tidal turbines in operation below the water's surface is limited, and physical inspections or maintenance must be performed on shore. This requires turbines to be shut down and lifted to the surface [2], which can be extremely costly and results in long periods of down time.

In addition, the high kinetic energy of tidal forces causes tidal turbines to experience high torques and thrust loading [3], inducing great mechanical stresses. In combination with marine corrosion and fouling, this can potentially cause accelerated wear and damage to occur to tidal turbine machinery. Faults should therefore be detected as soon as possible before damage propagates throughout machinery, resulting in costlier failures within the device.

Condition-based maintenance strategies [4], utilising intelligent condition monitoring systems, can therefore be greatly beneficial to the tidal power industry, allowing operators to perform maintenance before costly damage occurs, yet only when necessary to avoid unnecessary costs due to periodic shut downs for inspection.

Nevertheless, tidal turbines are a new application for condition monitoring. There is currently limited experience of the response of full scale turbines in operation, under both healthy and faulty conditions. This can be challenging for implementing condition monitoring systems, where many state of the art techniques require historical data for automatic detection, classification and prediction of faults.

This thesis therefore investigates approaches to applying state of the art condition monitoring to new applications with limited operational histories and experience of faulty behaviour, specifically tidal power. This work has involved the analysis of condition data from the Andritz Hydro Hammerfest HS1000 tidal turbine, a commercial scale tidal turbine in operation at the European Marine Energy Centre (EMEC) in Orkney, Scotland. Through this analysis, approaches for fault detection, diagnostics and prognostics have been developed, leading to a framework for applying intelligent condition monitoring in new applications.

1.2 Thesis Outline

Chapter 2 introduces tidal power, detailing tidal flow and the current state of tidal turbine design. The chapter then discusses the maintenance of tidal machinery, the need for condition-based maintenance strategies due to the difficulties introduced by their operating environment and the potential sources of damage within tidal turbines. The state of the art within condition monitoring as a whole, and then more specifically within tidal power, is then reviewed. The Andritz Hydro Hammerfest HS1000 tidal turbine is then introduced, discussing its design and operation, and the sources of data available for this work. Finally, an approach to condition monitoring for tidal turbine devices is detailed, following the Cross-Industry Standard Process for Data Mining (CRISP-DM) to perform anomaly detection, and then diagnostics and prognostics. Here, a framework for performing condition monitoring in new applications based on CRISP-DM is first introduced.

Chapter 3 details data mining through the CRISP-DM methodology, and how this can be utilised to construct models of normal behaviour for anomaly detection to detect faults within a condition monitoring system. This chapter describes how each stage of CRISP-DM was applied in this work to model responses from the HS1000 tidal turbine, discussing the use of data analysis techniques to reveal key relationships within datasets, the selection of appropriate modelling techniques given different forms of data, and how results can be interpreted by operators after models are deployed to assess the state of machine health.

Chapter 4 then discusses diagnostic modelling, where the specific fault mode can be classified following fault detection. The chapter first introduces vibration as a key parameter in rotating machine diagnostics and details the expected vibration characteristics for different fault modes of gearboxes, bearings and generators. The chapter then discusses feature-based methods of fault classification, where signal processing methods are used to extract characteristics from vibration signals unique to specific fault modes before classification. A deep learning approach to rotating machine fault classification is then presented, where deep neural networks are used to automatically extract fault characteristics without the use of tailored signal processing methods. These two approaches to fault classification are then applied through two case studies: identification of a change in the HS1000 generator response and identification of multiple faults within a gearbox test rig. Finally, the role of diagnostics within condition monitoring for new systems is discussed, expanding the CRISP-DM framework to include diagnostic modelling.

Chapter 5 details the use of prognostics within condition monitoring to predict the remaining useful life of components undergoing degradation. The requirements of different approaches to prognostics are then discussed, finding degradation-based methods, and in particular, model-based methods best suited to applying prognostics in new applications with limited operational histories. Degradation and model based methods are then examined in more detail, focussing on Bayesian filtering approaches (namely the Kalman Filter, Extended Kalman Filter, Unscented Kalman Filter and Particle Filter). The application of model-based prognostics in new systems is then discussed, including the selection of suitable degradation parameters and degradation models. An example of applying the Particle Filter for model-based prognostics is then given through a simulated gearbox degradation in the HS1000. The chapter concludes by further expanding the CRISP-DM framework to include prognostic modelling.

Chapter 6 finally concludes this thesis, summarising the work, its main findings and contributions, and detailing potential areas of further work that can be undertaken as a result.

1.3 Novel Contributions

The following points describe the novelty of the research undertaken as part of this thesis:

- Analysis of sensor data from a commercial scale operational tidal turbine data to identify key operating and condition parameters suitable for condition monitoring
- Use of change point analysis to identify control modes directly from data, without turbine design details or specifications
- Prototyping of anomaly detection models for detecting faults in tidal turbines from normal behaviour – suitable for condition monitoring within new applications with no prior experience of faults or fault data
- Development of a deep learning approach for diagnosis of rotating machine faults from vibration data
- An approach for implementing model-based prognostics in new systems with limited previous examples of failure
- Expansion of the CRISP-DM methodology for data mining to implement condition monitoring systems in new applications with limited operational histories

1.4 Related Publications

The following publications are related to this work and have been published prior to the submission of this thesis:

- G. S. Galloway, V. M. Catterson, C. Love and A. Robb, "Anomaly Detection Techniques for the Condition Monitoring of Tidal Turbines", *Annual Conference of the Prognostics and Health Management Society*, Fort Worth, TX, 29 Sep. – 2 Oct. 2014.
- G. S. Galloway, V. M. Catterson, T. Fay, A. Robb and C. Love, "Diagnosis of Tidal Turbine Vibration Data through Deep Neural Networks", *Third European Conference of the Prognostics and Health Management Society*, Bilbao, Spain, 5-7 Jul. 2016.

The following publications are related to this work and have been accepted for journal publication:

- G. S. Galloway, V. M. Catterson, C. Love, A. Robb and T. Fay, “Modeling and Interpretation of Tidal Turbine Vibration through Weighted Least Squares Regression”, *IEEE Trans. Systems, Man, and Cybernetics: Systems*, 2017.

Chapter 2

2 Tidal Power and Condition Based Maintenance

Tidal power has great potential worldwide to be a major contributing source of renewable energy. In 2009, the EU announced a Renewable Energy Directive [5], setting a target for 20% of energy generation to come from renewable resources by 2020. Within the UK, this target was adjusted to 15% [6]. Between 2005 and 2014, the UK has increased its renewable energy production from 1.5% to 7.0% [7], however there is some way to go to reach the 15% target.

Within the UK alone, tidal resources have the potential to yield up to 94 TWh of energy per year, with predictions showing tidal arrays have the potential to provide 9.55 GW capacity in the UK by 2020, with currently available turbine technology [1]. This accounts for approximately 31.3% of the UK’s 30.5 GW renewable energy capacity in 2015 [8]. It is therefore clear that tidal energy has the potential to provide a major contribution to renewable sources of energy.

In addition, tidal power holds a number of advantages over alternative renewable energy technologies (such as wind power) [9]:

- Predictability – tidal flow is highly predictable, as tidal cycles can be predicted months and years into the future
- High density of seawater – allows tidal turbines to have smaller blade diameters, reducing overall size and cost of devices for the same power output as a wind turbine
- High capacity factor – large amounts of energy can be extracted from tidal flows efficiently
- Slow moving rotor – reduces impact on surrounding environment
- Tide flows in (generally) only two directions – enables more efficient turbine design without need for rotating yaw of turbine nacelle.

However, the operating environment introduces unique challenges. The marine environment can introduce wear to tidal turbines through fouling (due to a build-up of soft algae or barnacles) or corrosion (due to the salinity of ocean water). In addition, the high loads of tidal flow may cause accelerated aging to turbines’ mechanical components [3]. Access to tidal turbines located on the sea bed for inspection or maintenance is difficult, and cannot be performed when the turbine is in operation. This leads to maintenance actions being extremely costly, where the turbine must be lifted from the sea bed and accessed on shore.

Condition based maintenance strategies [10] can therefore be of great benefit to the tidal turbine industry, where maintenance is only planned when necessary, based on analysis through state of the art monitoring techniques.

2.1 Tidal Flow

Tidal turbines convert kinetic energy from tidal flow to electrical energy. Tidal flow is a result of the Moon's and Sun's gravitational forces acting on the Earth, causing sea levels to rise and fall periodically in tidal cycles [11]. Aspects of tidal flow are therefore easily predictable, as tide height is directly related to the position of the Moon relative to the Earth.

The EMEC test site in Orkney experiences semidiurnal tidal flow [12], where two tidal cycles occur every lunar day. This leads to the HS1000 turbine experiencing approximately four tidal flows (two upstream and two downstream) each day. Figure 2.1 details a sample of modelled tide heights over time for the EMEC test site, showing the effect of semidiurnal tidal flow.

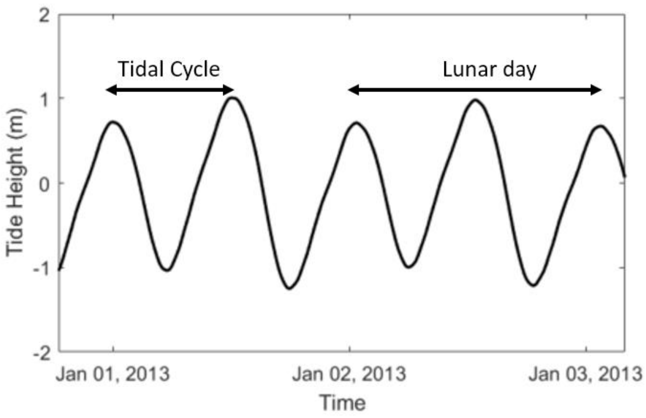
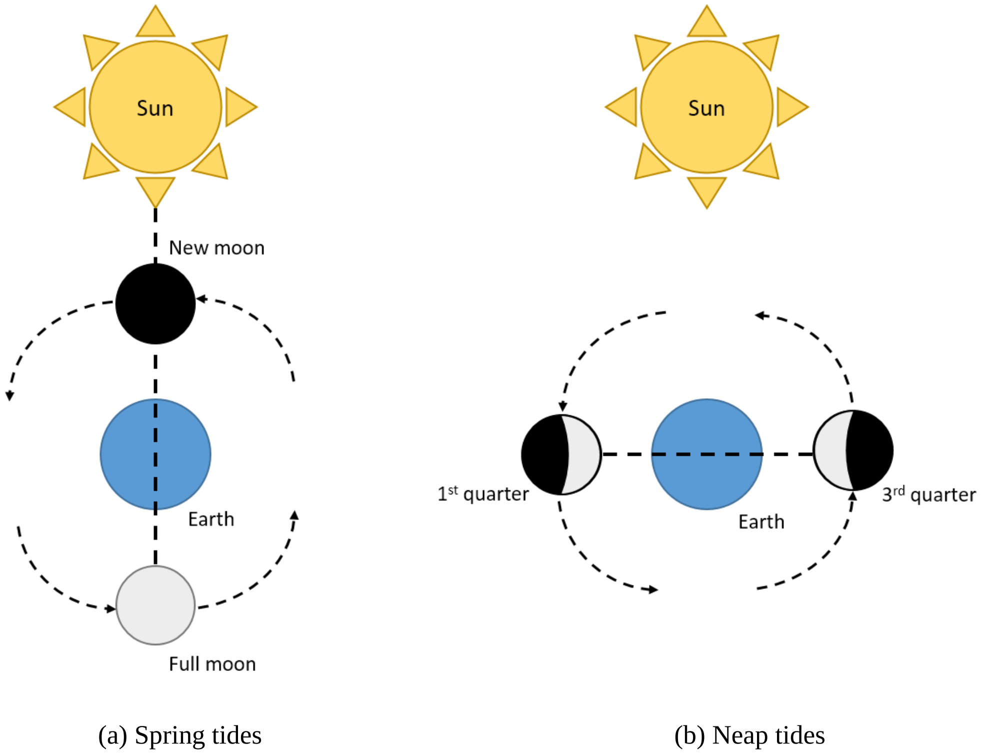


Figure 2.1 - Semidiurnal tidal flow, from modelled EMEC tidal data

Maximum and minimum tidal heights vary periodically over an approximate two-week cycle [13]. During new and full moons, the Sun, Moon and Earth are aligned. Additional gravitational forces due to the Sun are added to those of the Moon and cause variation in tide heights to be at their maximum. This is known as the spring tide, Figure 2.2 (a). Conversely, when the Moon and Sun are at 90 degrees in relation to the Earth, the Sun's gravitational forces counteract the Moon's gravitational forces and tide height variation is at a minimum. This is known as the neap tide, Figure 2.2 (b). Consequently, tidal flow rates are increased during spring tidal cycles, where the difference between high and low tide heights is greatest. Figure 2.3 shows the effect of spring and neap tides at the EMEC test site.



(a) Spring tides (b) Neap tides

Figure 2.2 - Phases of spring and neap tide

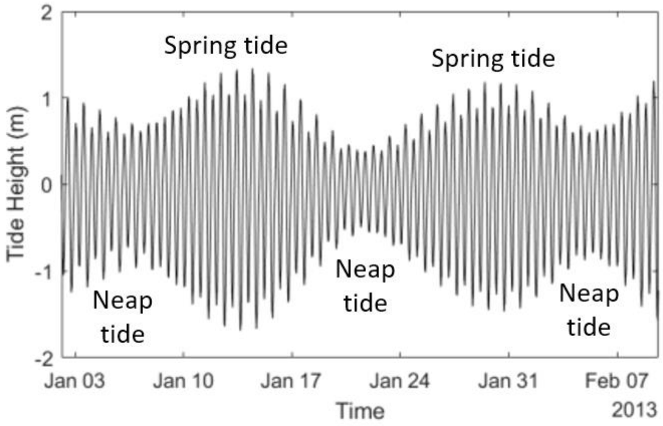


Figure 2.3 - Spring and neap tidal cycles, from modelled EMEC tidal data

2.2 State of the Art Tidal Turbine Design

Tidal power technology is in its infancy, and no clear turbine design has emerged as an industry standard for extracting energy from tidal flow. The state of the art in turbine design includes many horizontal and vertical axis solutions, some with major structural and operational variations [9].

A number of different tidal turbine designs have been proposed:

- Open rotor horizontal axis – Most similar to the traditional wind turbine design, with a horizontal rotor parallel to the direction of tidal flow
- Ducted horizontal axis – Contains a duct over the rotor to accelerate tidal flow, leading to greater rotational speeds and output power
- Multiple rotor horizontal axis – Contains multiple rotors on a single body, to capture additional tidal flow energy in a smaller form factor with reduced costs
- Transverse/vertical axis – Rotation of the device is perpendicular to tidal flow, with blades aligned vertically, aiming to improve structural efficiency

J. King and T. Tryfonas [14] and J. Zhang, et al., [15] have published reviews on the state of the art in tidal power systems currently in development from different manufacturers. This data has been updated for information currently available from each manufacturer and summarised in Table 2.1 (note this is not a comprehensive list of all tidal turbine manufacturers worldwide). This shows the majority of manufacturers have chosen variations of the horizontal axis design with power ratings commonly between 1 – 2.5 MW. Transverse axis turbines are in development, however these primarily consist of more conceptual designs with lower power ratings.

Drive train configurations differ between design types. The drive trains of open rotor and ducted horizontal axis turbines typically consist of an induction generator driven by the blades through a gearbox. Gearbox designs differ between manufacturers, but configurations commonly contain multiple stages with one or more planetary stages (as with Andritz Hydro Hammerfest, Atlantis and Alstom turbines).

Transverse/vertical axis devices are also typically driven by a gearbox-generator configuration [14], however generally generate energy at a lower power rating than horizontal axis devices (Table 2.1). Gearbox configurations therefore contain fewer stages.

Table 2.1 - Tidal turbine designs, manufacturers and proposed power ratings (updated from [14] and [15])

Turbine Type	Manufacturer	Generator Power Rating
Open rotor horizontal axis	Verdant Power LTD	35 kW
	Nautricity	500 kW
	Andritz Hydro Hammerfest	1 MW
	Atlantis	1.5 MW
	Free Flow Power	1 MW
	Swan Turbines	1 MW
	Alstom	1.4 MW
	Sabella	2 MW
	Aquamarine	2.5 MW
Ducted horizontal axis	Clean Current	65 kW
	UEKUS	300 kW
	Open-Hydro LTD	500 kW
	Lunar Energy	1 MW
	Atlantic Resources Corporation	2 MW
Multiple rotor horizontal axis	Statkraft	1 MW
	Sustainable Marine Energy LTD	1 MW
	Teamwork Technology BV	1 MW
	Voith Siemens	1 MW
	Marine Current Turbines	1.2 MW
	Scot Renewables	2 MW
	Aquamarine	2.4 MW
	Tidal Stream Energy	10 MW
	Bluewater Energy Services	n/a
Transverse/vertical axis	Ocean Renewable Power	22 kW
	Ponte di Archimede	13 kW
	Blue Energy	200 kW

2.3 Tidal Turbine Maintenance

The underwater marine environment poses a great challenge for the operation and maintenance of tidal power devices. Strong kinetic forces from turbulent tidal flow lead to tidal turbines experiencing extremely high loads [3]. This could potentially cause accelerated wear and damage to the turbine and its internal rotating components. Other factors such as marine fouling and corrosion may also lead to loss of performance and accelerated damage, primarily affecting the turbine's external components such as blades, nacelle and support structure [3].

While components have been designed and manufactured to operate in such conditions (including the use of specialised coatings to resist marine fouling and corrosion, and materials to withstand high loads) potential fault causes of tidal turbines are not well understood in practice, with the degradation rates of such faults remaining uncertain [16].

Access to tidal turbines while in their operational environment for routine inspection and maintenance is extremely challenging. Assessing the external health of the machine may be possible through remote surveillance equipment. Divers may also be used to inspect the exterior of the machine and perform maintenance actions such as scouring to remove marine fouling, but this may be dangerous and can only be performed during favourable weather conditions [17].

Access to the internal components of the turbine for inspection or maintenance is not possible in a submerged environment. Turbines must therefore be designed to be retractable from the seabed, allowing internal component maintenance activities to be performed onshore. This requires costly specialist marine equipment, e.g. a ship fitted with a large winch or crane, and results in prolonged periods of machine downtime. This process is also highly dependent on weather conditions, as the lifting operation cannot be performed safely when waves and turbulence are above certain levels [18].

2.3.1 Condition Based Maintenance

Maintenance strategies broadly fall into three main categories [19] [20]:

1. Preventative maintenance
2. Corrective maintenance
3. Condition based maintenance

Preventative maintenance strategies aim to prevent sustained damage to machinery by scheduling maintenance periodically in time [21]. In this strategy, faults can be identified at an early stage through inspection and damage to equipment is kept minimal. However, in order to increase the reliability of the machine and reduce costs due to failures occurring in operation, maintenance must be performed before the mean time to failure [19]. This category of maintenance therefore involves the highest maintenance costs, and will be particularly high within tidal power due to the difficulty in accessing machinery in the marine environment.

Corrective maintenance strategies involve only performing maintenance when failure occurs [21]. Such a strategy may reduce costs associated with unnecessary inspections and maintenance, but greatly increases the risk of more catastrophic failure occurring within the machine, leading to higher costs due to component repair and replacements [20]. This issue is particularly prevalent for tidal power, as faults occurring in a tidal turbine's components during operation may go unnoticed without a sophisticated monitoring system.

Condition based maintenance is an evolution of traditional preventative and corrective maintenance strategies, where maintenance is performed only when necessary, based on an estimation of the health of a machine's components [22]. This is illustrated in Figure 2.4, adapted from [19], where condition based maintenance aims to minimise total costs due to both maintenance and repair actions. Modern advances in sensor technology and data analytics allow periodic inspections to be supplemented by condition monitoring systems, where the health state of components can be inferred remotely from sensor data and faults can be detected at an early stage.

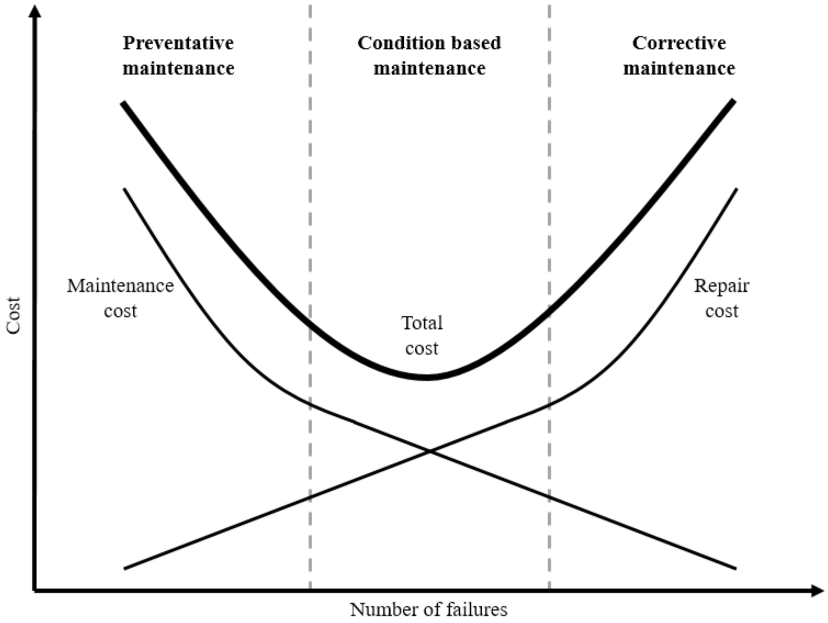


Figure 2.4 - Cost against number of failures for different maintenance strategies, adapted from [19]

Condition based maintenance strategies can therefore be extremely beneficial within the field of tidal power, reducing maintenance and repair costs and improving the reliability and availability of machinery. However, an effective condition based maintenance strategy heavily relies upon an accurate condition monitoring system. While limited remote inspection of the external of the turbine is possible to assess a turbine's health (performed by divers or surveillance equipment), detailed inspection of a turbine's internal components is not possible during operation. Effective condition monitoring techniques can provide operators with an accurate remote assessment of the health of internal components, detecting faults before failure occurs.

2.3.2 Sources of Damage Within Tidal Turbines

The design of horizontal axis tidal turbines is very similar to that of most wind turbines. It may therefore be expected that such tidal turbines will experience some similar fault modes. However, despite similarities in turbine design with wind power devices, the operating environment is vastly different. Water is over 800 times denser than air and, despite slower flow rates (around 3 m/s compared to around 15 m/s for offshore wind), tidal flow has a much higher kinetic energy compared to wind flow [3]. This causes tidal turbines to operate with higher torque and thrust loading, inducing increased stress on the machine, particularly on the low speed stages of the drive train. Additionally, the marine environment provides other complications, such as corrosion and interaction with plant and animal life [3].

P. Beaujean, et al [23], discuss reliability concerns of tidal turbines in contrast to other power generation technologies. It is stated that a main reliability concern is the fouling of mechanical components due to the corrosive environment of sea water, an issue unique to submerged tidal turbines. Another stated concern is underwater turbulence. Waves on the surface can create uneven loading between the top and bottom of the turbine, which may lead to additional wear on the blades, rotor and supporting components such as bearings and the gearbox.

P. Prickett, et al [10], use failure modes and effects analysis (FMEA) to analyse potential failure modes within tidal turbines. Key fault modes identified include corrosion and impacts to the rotor blade assembly, wear and fatigue to the main drive shaft due to high loads, and imbalance and misalignment of the generator drive train.

L. Chen and W. Lam [16] list various aspects of survivability and remedial actions for tidal turbines. Primary sources of damage include extreme weather, seabed scour, fatigue failure, corrosion/erosion and marine fouling. Maintenance actions include manual cleaning of fouling, reapplication of coatings and paint and nacelle re-sealing to deal with sources of damage unique to the marine environment. It is also stated that monitoring of internal components (such as the gearbox, generator, etc.) is essential to allow early detection of potential causes of failure and keep maintenance costs affordable.

Despite the differences in the operating environment, a useful statistic to infer from wind power is the distribution of downtime for failures of similar mechanical components, where maintenance actions for such failures will be similar. Figure 2.5 shows the distribution of downtime for wind turbines as part of the Wind Measurement and Evaluation Program (WMEP), adapted from [24], a survey conducted to study the long-term reliability of wind turbines. This survey contained data from a number of different wind turbine designs and locations across Germany, with approximately 45% deployed offshore. This shows that faults within the turbines' primary internal rotating components (the gearbox, generator and drive train) contribute the largest percentage of downtime.

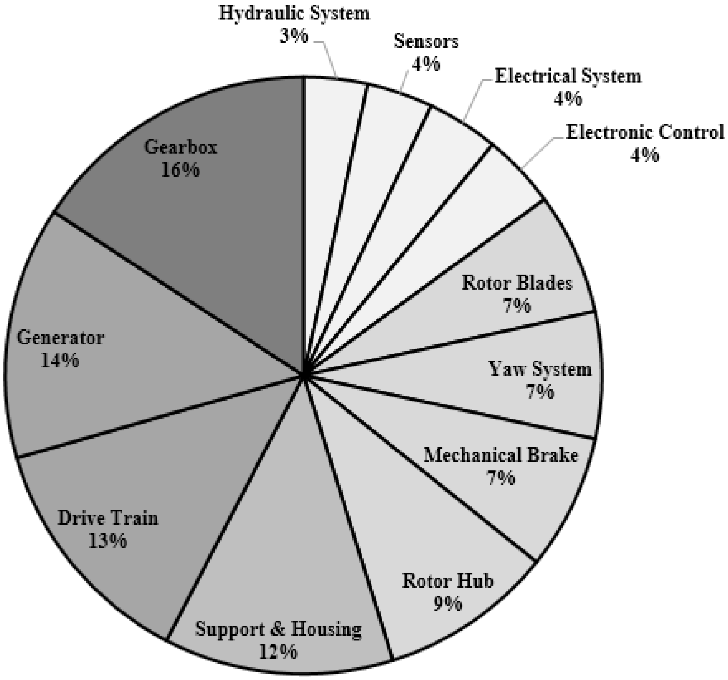


Figure 2.5 - WMEP wind turbine downtime distribution, adapted from [24]

Due to these factors, the main focus of this research has been the application of condition monitoring techniques to a tidal turbine's internal rotating components: the gearbox, generator, shafts and supporting bearings. Effective monitoring of these components was viewed as critical to a condition based maintenance strategy, as no reliable method of internal inspection exists during periods of operation, and repair actions for these components can result in the longest periods of downtime.

2.3.3 Current State of the Art in Condition Monitoring

Effective condition monitoring techniques are an essential part of a condition based maintenance strategy. While the aim of such methods is to detect faults occurring within machinery at an early stage, state of the art condition monitoring techniques can provide operators with an increasingly detailed assessment of the health of a machine and its individual components.

Within this research, condition monitoring techniques have been split into three categories: anomaly detection, diagnostics and prognostics.

Anomaly detection techniques are used to distinguish changes in operational behaviour, indicative of a fault or failure within the system, without the need for prior knowledge of component failure [25]. Such techniques can be performed by identifying deviations from models of normal behaviour, and therefore require measured data from the healthy system over a full range of operating conditions. In addition to recognising changes in behaviour, anomaly detection may allow operators to localise a detected fault to an individual component of the machine, or infer the level of damage that has occurred. Still, such techniques do not provide specific information about the nature of a fault.

Diagnostic techniques, however, allow operators to detect and quantify specific fault modes for individual components of a machine [26]. This is commonly performed for rotating machines through high resolution data, as opposed to data acquired at a low sample rate, e.g. SCADA data, which can lack the detailed information content required to make accurate fault diagnoses [27]. Research in this area has shown a focus on the development of signal processing techniques applied to high resolution vibration data [28] to best isolate known characteristic signals associated with specific faults (this is known as feature-extraction [29]). This requires prior knowledge of the response of components under normal and faulty conditions, inferred from expert knowledge or the analysis of fault data.

Finally, prognostics techniques provide an estimation of when a component will fail given a specific fault mode [26]. This is known as the ‘remaining useful life’ (RUL) of a component, and allows an operator to infer how long a machine can remain active before the component fails and a maintenance action is required. Prognostic techniques require prior knowledge of the response of the system under both normal and faulty conditions, as well as an understanding of the degradation process of components experiencing wear and faults. This degradation process can be modelled mathematically, or inferred through historical data [30].

2.3.4 State of the Art Condition Monitoring for Tidal Turbines

As tidal power is still an emerging power generation technology, it is not yet an established application within condition monitoring. Understanding of the failure modes of standard components of tidal turbines is known in theory, yet there is a limited understanding of the response of tidal turbines in operation [16]. Condition monitoring methods are recently becoming more established within the wind energy field [28] [31] [32], and due to the similarity in design between wind and tidal turbines, some developed techniques may be transferable. However, due to a vastly different operating environment and a lack of historical fault data, different approaches should be examined.

P. Caselitz and J. Giebhardt [33] discuss condition monitoring for tidal turbines, detailing the required hardware and potential fault prediction algorithms. The following parameters were identified as useful indicators of faults:

- Acceleration and vibration within bearings and gear wheels
- Strain across blades
- Displacement on slow rotating shafts and bearings
- High frequency generator phase currents
- Gearbox oil quality

Vibration analysis is then identified as a technique that may be suitable within this environment, using the Fast Fourier Transform (FFT) to identify bearing and gearbox faults [33]. Monitoring of the overall performance of the turbine, examining parameters such as output power with reference curves generated from the first months of operation, was also identified as a useful indicator of health.

As well as vibration analysis, analysis of electrical was also identified as suitable for condition monitoring. Electrical signals, particularly high frequency phase and current data, can be used to detect faults within a turbine's generator (e.g. rotor or stator winding faults) [34].

R. Wald, et al [35], review a number of techniques potentially suitable for condition monitoring and prognostics of tidal turbines. The authors state that despite similarities in design with wind turbines, a different approach is required due to the differences in operating environment. Feed forward multi-layer neural networks, radial basis function neural networks and adaptive neuro-fuzzy inference techniques are reviewed. However, it is stated that these methods require data from healthy and failure states of the system, which is not readily available for tidal power.

Practical implementations of condition monitoring techniques for tidal turbines have been published, namely research from Florida Atlantic University. A 20 kW dynamometer has been constructed to simulate the drive train of a tidal turbine [36]. The test rig contains a motor to simulate rotating blades, two gearboxes and a generator. Condition monitoring techniques applied to this dynamometer have focussed on locating bearing faults through vibration analysis of accelerometer data. A number of methods within vibration monitoring have been explored for fault detection and diagnosis:

- Combination of Wavelet and Fourier Transform [37]
- Short time Wavelet analysis [38]
- Machine learning classification of Wavelet features [39], [40]

Data fusion techniques have also been explored to combine multiple features extracted from wavelet analysis. Fusion was applied at the data level [41], feature level [42] and decision level [43] to improve the accuracy of classification algorithms for bearing fault diagnosis.

As a result, a 'Smart Vibration Monitoring System' was developed [44]. The system combines a number of vibration monitoring techniques, including FFT, Short Time Fourier Transform, Wavelet, Cepstrum and time-domain measures (Kurtosis, RMS, peak level), to detect bearing and gear mesh faults.

2.4 HS1000 Tidal Turbine

Data for this project was sourced from the Andritz Hydro Hammerfest HS1000 tidal turbine, a full commercial scale prototype tidal turbine. This turbine was operational at the European Marine Energy Centre (EMEC) in Orkney, Scotland, a test site used by ocean energy developers to test state of the art technology fully connected to the grid in a typical marine location.

The HS1000 turbine has an open rotor, horizontal axis design, shown in Figure 2.6. The turbine body is held stationary in depth and direction by a three-legged support structure attached to the sea bed. The turbine's blades rotate bi-directionally (both clockwise and anti-clockwise) with variable speed in response to upstream and downstream tidal flows.

Figure 2.6 - Andritz Hydro Hammerfest HS1000 tidal turbine (courtesy of Andritz Hydro Hammerfest)

2.4.1 HS1000 Design

Figure 2.7 shows the internal design of the HS1000 tidal turbine. The turbine's structure is similar to many modern wind turbines, where three blades drive a shaft connected to a generator through a gearbox, converting tidal flow into electrical power.

This turbine contains a three-stage gearbox manufactured by David Brown Gear Systems, increasing rotation speed from the blades to the generator by a factor of 98.43 up to speeds exceeding 1000 RPM. The first stage is a planetary gear stage, shown in Figure 2.8, where the ring gear is fixed in position, the planet carrier is connected to the input shaft (driven by the blades) and the sun gear is connected to the input of the next gear stage. This planetary gear configuration allows extremely high loads from tidal forces to be distributed between ring, planet and sun gears, where no single gear pairing takes a full load [45]. The second and final stages of the gearbox are both parallel axis gear stages, further increasing the rotation speed at each stage.

The HS1000’s generator is a water-cooled 6600 V doubly-fed induction generator (DFIG), manufactured by ATB Morley. A DFIG contains a frequency convertor connected between the generator’s rotor windings and the grid supply, allowing rotor frequency to vary from the frequency of the grid [45]. This allows the rotation speed of the generator to be varied through torque control.

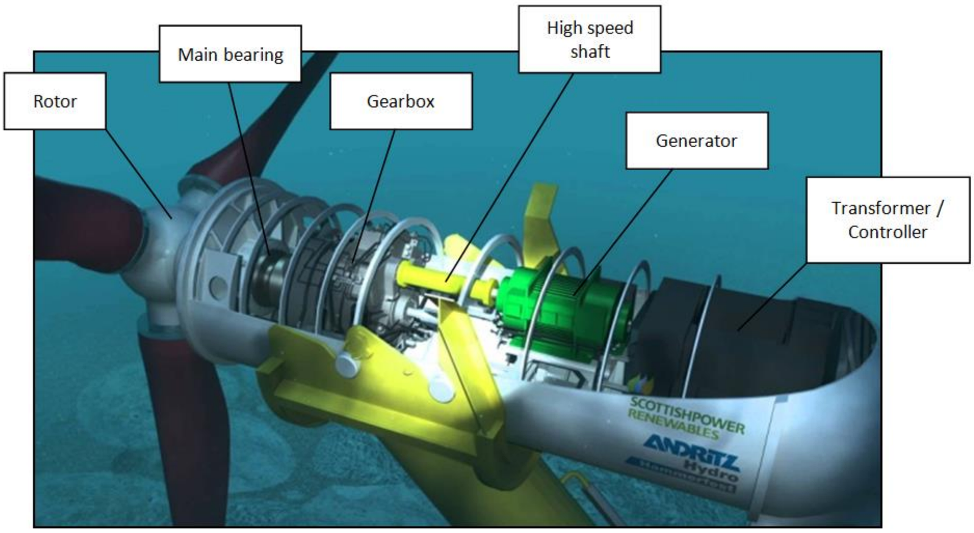


Figure 2.7 - Internal design of the HS1000 tidal turbine (courtesy of Andritz Hydro Hammerfest)

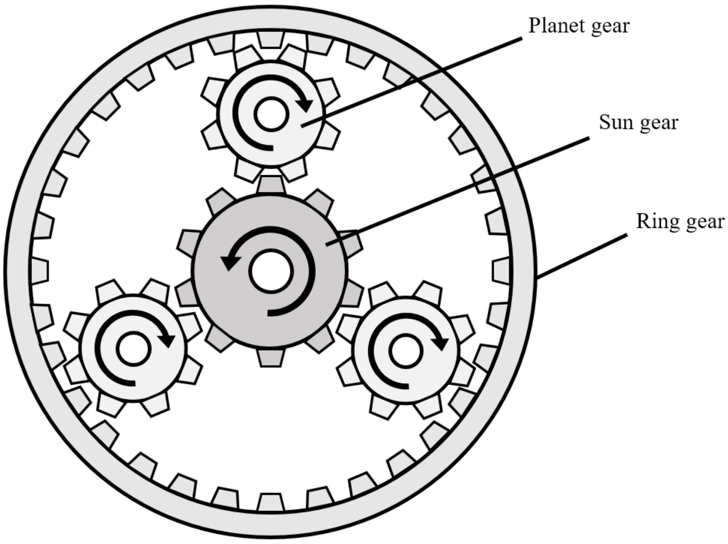


Figure 2.8 - Planetary gear stage

2.4.2 Turbine Operation

The HS1000 tidal turbine was active over two separate deployment periods at the EMEC, including operation between December 2011 to January 2013 and August 2013 to February 2014. During each deployment period, the HS1000 performed continuous automated operation and experienced a range of tidal flows, while providing electrical energy to the UK grid. Between operating periods, the turbine was shut down and components were inspected by supply chain partners, where no faults to internal rotating components were identified.

In operation, the HS1000 turbine was regulated through a control scheme, utilising blade pitching and generator torque control to manage the rotation speed of the machine and output power in response to different levels of tidal flow.

2.4.2.1 HS1000 Control Scheme

The HS1000 tidal turbine's control scheme controls the rotation speed of the machine in relation to different tidal flow rates. This allows the turbine to maintain efficient levels of output power while operating within specified limits. Control is performed through a combination of blade pitching and generator torque control.

Blade pitch control involves rotating the turbine's blades to catch more or less of the oncoming tidal flow, similar to pitch control within wind turbines [45]. Increasing the angle of the blades to catch more of the oncoming tidal flow (the angle of attack) increases the drag against the blades. This captures more of the tidal force and increases loading on the machine, causing blades to rotate at a faster rate. Conversely, loading and rotation speed can be reduced under high tidal flow rate by decreasing the blades' angle of attack and reducing drag.

Pitch control is primarily performed to reduce the speed of the machine at high tidal flow rates to maintain steady levels of rated power and stop the machine operating above its specified limits [45]. Pitching is also performed during shutdown periods, where the blades' angle of attack is fully decreased (known as feathering) to minimise loading on the machine. A mechanical brake within the turbine is then applied to stop any rotation. The turbine is shut down manually during periods of adverse weather (e.g. high winds, where waves cause excessive tidal turbulence).

Torque control also controls the rotation speed of the machine, and is performed within the turbine's doubly-fed induction generator, again similar to generator torque control within a wind turbine [45]. Torque is varied within this configuration of generator using a variable frequency drive connected between the generator's rotor and the grid supply. The electrical frequency of the rotor's windings is varied independently from the constant frequency (50 Hz) of the stator windings allowing for variation in generator torque. Decreasing generator torque below the torque harnessed by the turbine's blades causes the rotor to accelerate, and increasing the generator torque above torque harnessed by the blades causes the rotor to decelerate. This is performed when the tidal flow rate is below rated value, while pitch control keeps blades at a constant and most optimal position, allowing generator rotation speed to increase as tidal flow rate increases, capturing more power. Above rated tidal flow rates, torque control is used with pitch control to limit the rotation speed and reduce loading on the machine.

The HS1000's control scheme differs slightly between upstream and downstream tidal flows. During upstream tidal flow, where the tide flows across the turbine from blades to the rear of the turbine, the machine operates up to its rated power of 1 MW, rotating at speeds exceeding 1000 RPM within the generator. During downstream tidal flow however, where the direction of flow starts from the rear of the turbine, blades are pitched more during high flow rates to reduce the tidal loading on the machine. This is enforced to ensure the turbine is not lifted from its support structure. Therefore, during downstream flow, the maximum rotation speed of the turbine is limited to around 800 RPM and the turbine does not reach maximum rated power.

2.4.3 Data Sources

The HS1000 was fitted with a wide variety of different sensors, measuring a range of different operational and condition parameters during the turbine's operation.

Data used in this work was mostly obtained through two different systems: low frequency programmable logic controller (PLC) data, and high resolution accelerometer data.

2.4.3.1 Low Frequency PLC Data

Low frequency PLC data were recorded from a variety of sensors within the turbine for monitoring and control purposes. This data included numerous vibration velocity, displacement and temperature readings from different components, and operational data detailing the rotation speed of the machine and output power values.

This data was made available from the second deployment of the HS1000, for operation between September 2013 and February 2014. The data were recorded at an inconsistent sampling rate, varying up to a maximum of 10 Hz and were accessed for this project through the VantagePoint platform. This platform allowed multiple data parameters to be visualised over custom time intervals. The tool offered no form of data analysis beyond summarising data through maximum and minimum values. The data were therefore exported for further analysis using additional software tools, e.g. MATLAB.

2.4.3.2 High Resolution Accelerometer Data

High resolution accelerometer data was recoded from a CX1 accelerometer sensor, located on the nacelle of the turbine adjacent to the generator.

This sensor gave measurements of X, Y and Z axis acceleration (g) at a 2 kHz sample rate, as well as temperature (°C) at 1 Hz, and X and Y tilt (degrees) at 20 Hz.

Samples of high resolution vibration data were available for both deployment periods of the HS1000. Data were recorded during manually triggered sampling periods. A number of samples had been recorded for analysis over both deployments, detailing the vibration response during different phases of turbine operation.

The accelerometer data were accessed through the Sensview platform, allowing data from each sampling period to be examined, where both the time and frequency domain representations of vibration data could be displayed. Data was exported using this tool for further analysis in additional software tools, e.g. MATLAB.

2.4.3.3 Approach to Condition Monitoring for Tidal Power

Considering the current state of the art, the role of this research has been to study how such condition monitoring techniques can be applied in practice, to industrial scale tidal turbine devices operating within real marine environments and experiencing full tidal loading, but with limited operational history. As previously stated, this research has focussed on the monitoring of a turbine's internal rotating components (such as the gearbox, generator, shafts and bearings) using vibration data.

This first involved an in-depth analysis of recorded condition data from the HS1000 tidal turbine in operation through data mining and the CRISP-DM process model [46], detailing how the response of an industrial scale tidal turbine changes under different loading conditions, driven by phases of the turbine's control scheme. These responses were then modelled to define the normal response of the turbine, from which anomaly detection can be performed.

Secondly, state of the art diagnostic modelling techniques were investigated to identify a specific fault mode within the turbine's generator during operation from high resolution accelerometer data. This involved a comparison between traditional feature-based classification approaches and deep learning methods.

Finally, state of the art prognostics techniques were studied, focussing on model-based approaches and how these methods can be adapted for new systems like tidal power, with limited operational history and previous examples of failure.

As a result, an adaptable framework is proposed for the design of intelligent condition systems within new application areas, Figure 2.9. This framework is based on CRISP-DM and integrates the processes of all three categories of condition monitoring techniques (anomaly detection, diagnostics and prognostics) into a single system as more fault data becomes available. The framework takes into consideration the limitations of working with newly commissioned machines, and is flexible to accommodate the introduction of new modes, and updating of current models, as new data is recorded.

The first column in Figure 2.9 details an initial process to be performed on condition data, identifying key relationships within the data and performing modelling for anomaly detection. This work is detailed in Chapter 3. Secondly, following the identification of fault data through anomaly detection, diagnostic modelling can be performed, as in the second column in Figure 2.9, and as detailed in Chapter 4. Finally, if data is showing a suitable degradation trend, prognostic modelling can be performed, as detailed in Chapter 5.

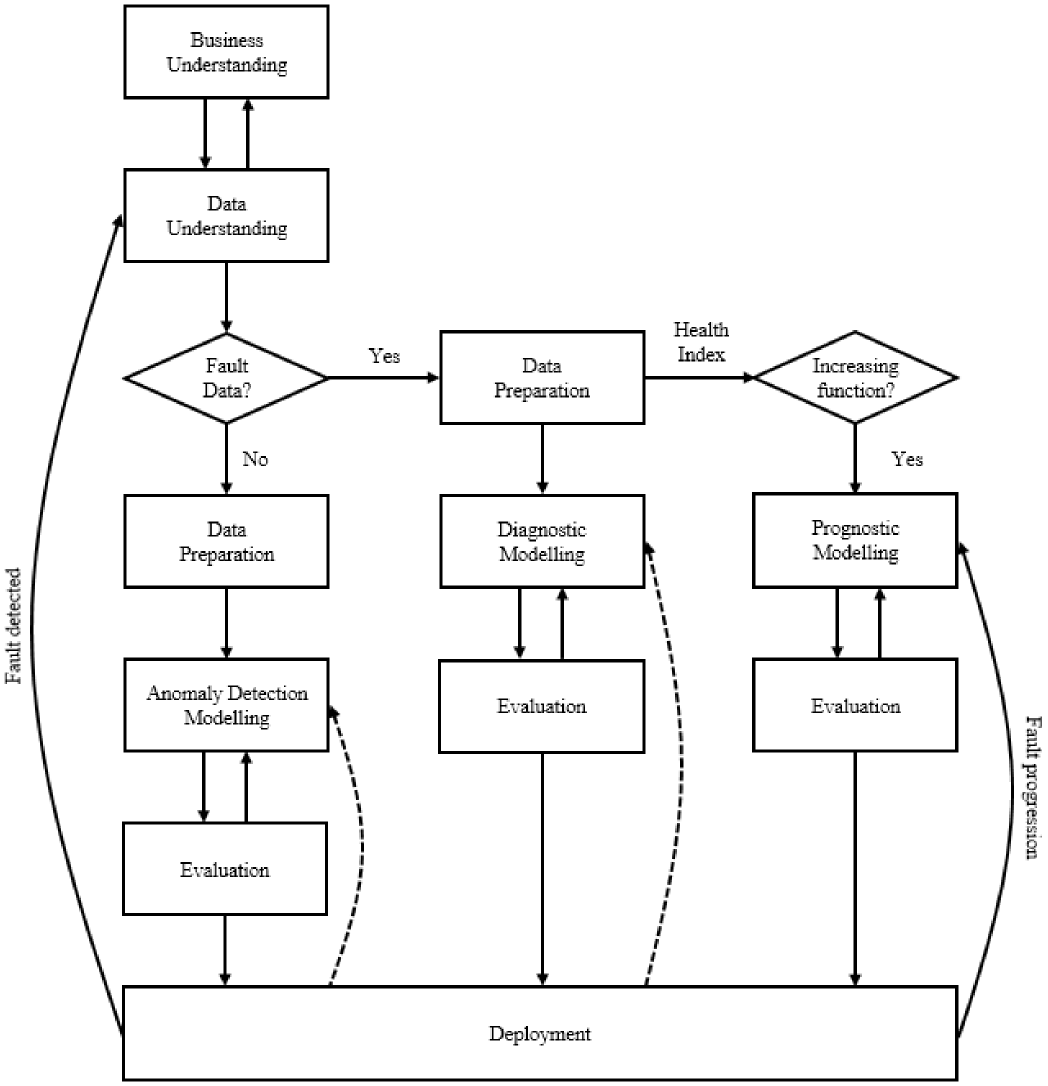


Figure 2.9 - Process for designing intelligent condition monitoring systems in new application areas, based on CRISP-DM [46]

Chapter 3

3 Anomaly Detection through Data Mining

The development and application of condition monitoring methods to new systems, with limited operational history and experience of faults in the field, (e.g. tidal power) can be challenging. Data mining, however, is a good starting point for condition monitoring within such systems.

Data mining is the process of knowledge and pattern discovery within datasets, turning low-level data into more usable forms, i.e. modelling [47]. In this case, data from periods of operation where a limited number of faults have occurred (or none at all) within a system can be analysed through data mining and used to construct models of normal behaviour. Anomaly detection can then be performed using such models, identifying changes in the response of components' condition parameters (such as vibrations, temperatures, etc.) in relation to the operation of the system, allowing faults to be detected at an early stage without prior experience of faults occurring in practice.

In this work, data mining was performed by following the CRISP-DM process model [46], utilising datasets of low resolution data from the HS1000 tidal turbine, containing multiple operational and condition parameters recorded from the turbine's gearbox, generator and high speed bearings.

This chapter first describes data mining, and the CRISP-DM methodology. The chapter then describes how each stage of the CRISP-DM methodology was used within this work, detailing how data mining was used to identify key trends and relationships within low resolution data to link measured condition parameters of components (e.g. vibration, temperature, etc.) to measured operational parameters (e.g. rotation speed, output power, etc.).

From here, anomaly detection modelling techniques were examined and utilised to define the normal behaviour of the turbine. After training, these models were tested on data from a range of operating conditions and were capable of identifying various anomalous events over the course of the turbine's operation: including a change in the control scheme, a pitch fault, a shutdown due to high ocean waves, and simulated gearbox faults. However, no actual damage to the turbine's internal components was inferred through this analysis on the available data.

3.1 Data Mining

Data mining is the analysis of large data sets for knowledge discovery, involving the use of processing techniques, statistical analysis, machine learning and visualisation methods to extract patterns and relationships hidden within data parameters [48]. Data mining has been commonly used by banking and marketing firms for knowledge discovery from financial data, and also within the medical field applied to vast amounts of patient records for improved diagnosis and prediction [49].

Within this study, data mining was used to discover trends and relationships between parameters within low resolution datasets from the HS1000 tidal turbine. A modelling stage then defines the expected response of the turbine over its typical range of operating conditions. By comparing live turbine data to these models, anomaly detection is used to indicate a change in the response of the system, indicating the possible onset of a fault.

3.1.1 CRISP-DM

The CRISP-DM (Cross-Industry Standard Process for Data Mining) process model was utilised for this study. This model manages the data mining process as six key stages: business understanding, data understanding, data preparation, modelling, evaluation, and deployment [46]. These stages are shown in Figure 3.1.

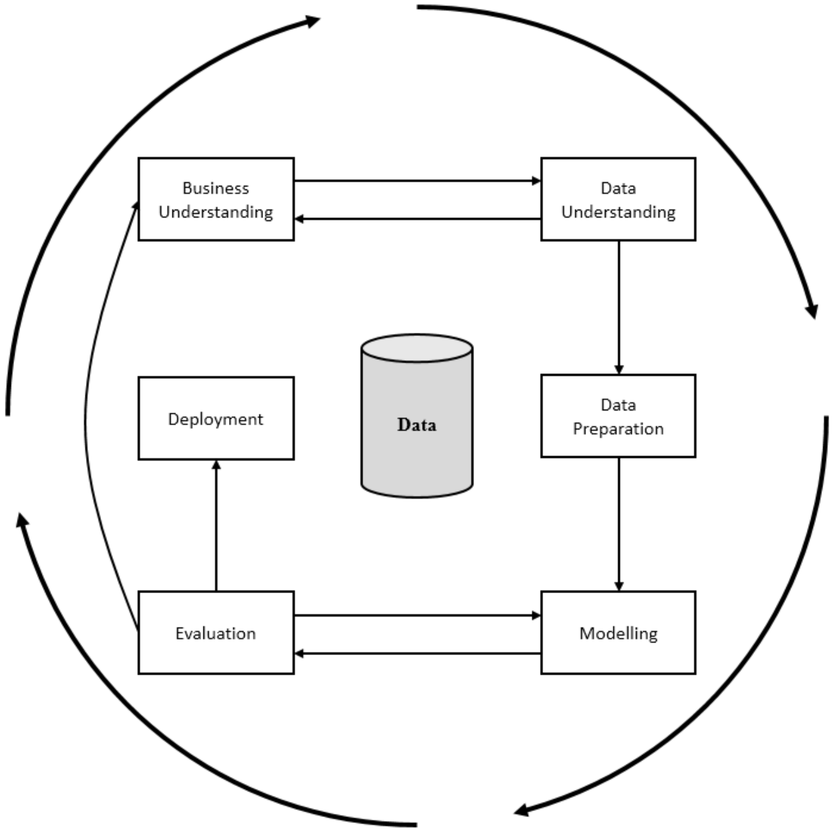


Figure 3.1 - The CRISP-DM process model for data mining adapted from [46]

Each phase of the CRISP-DM process model involves the following:

- Business Understanding – define the objectives and determine the goals of data mining
- Data Understanding – includes exploration of data through visualisation or data analysis techniques to identify key patterns or relationships
- Data Preparation – clean and organise the data before modelling
- Modelling – define the key patterns and relationships within the data using appropriate modelling techniques
- Evaluation – test the performance of models
- Deployment – use models and knowledge gained from the data to complete the defined objectives.

Alternatives to CRISP-DM include SEMMA (Sample, Explore, Modify, Model, and Access) or KDD (Knowledge Discovery in Databases) [47]. However, although many stages of each framework are similar, CRISP-DM allows for focus on understanding the data from both a business and data perspective.

Each subsequent section of this chapter discusses in detail how each phase of CRISP-DM was performed in this work, to understand the normal operation of the HS1000 tidal turbine to form and evaluate anomaly detection models from low resolution condition data.

3.2 Business Understanding

The first stage of the CRISP-DM methodology is ‘Business Understanding’. This stage involves initially defining the objectives and requirements of the work from a business standpoint. This definition is then used to define the data mining problem, and to determine the goals of the data mining work [46].

Within this work, the business requirements were to use the available condition data from the HS1000 to construct models that can be used within an intelligent condition monitoring system, capable of detecting faults within the turbine at an early stage before failure.

As discussed in chapter 2, this process would first involve performing anomaly detection, where it can be assumed that data from newly commissioned machinery is initially representative of normal behaviour.

As a data mining problem, this would involve identifying key trends and relationships within the available condition data, understanding the modes of turbine operation and relating this to measured condition parameters. These relationships would then be modelled to define the normal condition response of the turbine over its operation, selecting the most appropriate modelling technique for each condition parameter. Anomaly detection could then be performed, identifying changes in the normal response from each model indicative of a change in operation or the development of a fault within the turbine.

The data mining goals, developed after discussion with engineers at Andritz Hydro Hammerfest, were therefore defined as follows:

1. Explore the available data and identify key links and relationships between parameters that monitor the condition of components (such as vibration, displacement and temperature) and parameters that describe the turbine’s state of operation (such as tidal flow rate, machine rotation speed and output power)
2. Use properties of the data to identify distinct modes of turbine operation in relation to both operational and condition parameters
3. Model the normal response of turbine components using appropriate anomaly detection techniques
4. Identify changes in machine response and detect faulty behaviour using models of normal behaviour

3.3 Data Understanding

Following the definition of data mining goals in the business understanding stage, the data understanding stage of CRISP-DM involves an exploration of the available data, discovering the properties of the data before the modelling stage [46].

In this work, the data understanding stage involved examining the available data parameters, then the use of data analysis techniques to complete the first two data mining goals (defined in section 3.2).

This data analysis initially involved the use of correlation and principal component analysis to find relationships within the data, identifying key driving and response parameters. These parameters were then explored further using change point analysis, exposing operating modes that formed clusters within the data. This was found to be a suitable data-driven technique for modelling the stages of the turbine's control scheme without prior knowledge of the system or its design.

3.3.1 Data Parameters

Table 3.1 and Table 3.2 list the data parameters available for this study. Datasets contained low resolution data recorded over periods of turbine operation between September 2013 and February 2014. During this period, the turbine experienced a range of tidal conditions, including spring and neap tides during calm and adverse weather conditions. No faults to internal rotating components were reported during this period. The turbine has currently been decommissioned and is awaiting further inspections.

Parameters relating to the condition of the main internal rotating components are listed in Table 3.1, including vibration velocity, displacement and temperature readings for the generator, gearbox and high speed shaft supporting bearings.

Table 3.2 lists parameters describing the operation of the HS1000 tidal turbine including generator rotation speed, output power and tidal flow rate. The tidal flow rate parameter was generated from an interpolated EMEC model of expected tidal height, provided by Andritz Hydro Hammerfest. This parameter gives an approximation of the influence of periodic tidal flow on the turbine, however does not effectively model the influence of tidal turbulence caused by waves, wind and other adverse weather conditions. Measured rotation speed and output power parameters may therefore differ slightly in relation to the modelled tidal flow rate.

Figure 3.2, Figure 3.3 and Figure 3.4 show the response of these condition parameters over a typical tidal cycle. Above tidal flow rates of approximately 1.4 m/s the turbine is accelerated up to a rotation speed of around 800 RPM. Above tidal flow rates of approximately 2 m/s the turbine increases its rotation speed up to around 1000 RPM. As tidal flow rate decreases, the turbine decelerates back to 800 RPM, and finally deactivates as tidal flow falls to 1.4 m/s. The level of output power increases and decreases in relation to both tidal flow rate and generator rotation speed.

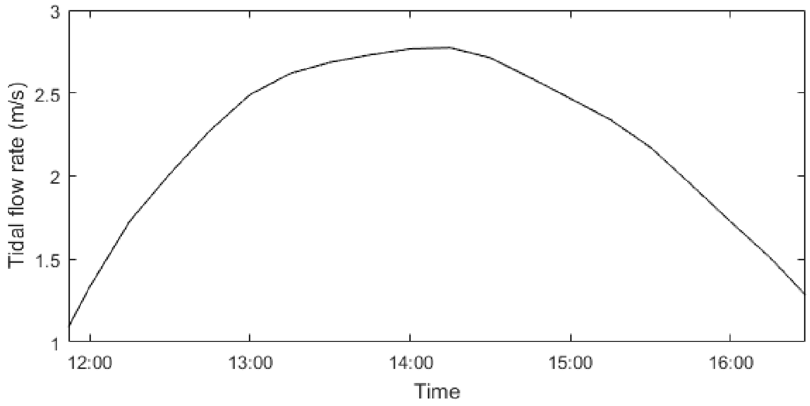


Figure 3.2 - Modelled EMEC tidal flow rate over time for one tidal cycle (16th Dec 2013)

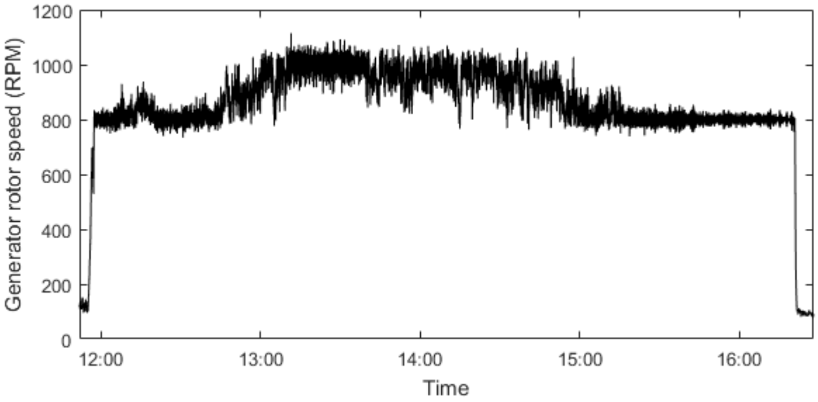


Figure 3.3 - Measured generator rotor speed over one tidal cycle (16th Dec 2013)

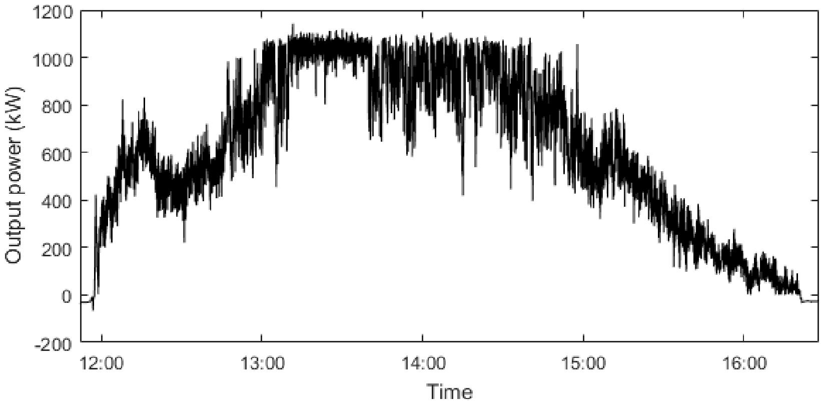


Figure 3.4 - Measured output power over one tidal cycle (16th Dec 2013)

Table 3.1 - Available HS1000 low resolution condition parameters

#	Sensor location	Parameter	Units
1	Generator – Non-drive end	Vibration velocity (x-axis)	mm/s
2		Vibration velocity (y-axis)	mm/s
3		Vibration velocity (z-axis)	mm/s
4	Generator – Drive end	Vibration velocity (x-axis)	mm/s
5		Vibration velocity (y-axis)	mm/s
6		Vibration velocity (z-axis)	mm/s
7	Gearbox – Output side	Vibration velocity (x-axis)	mm/s
8		Vibration velocity (y-axis)	mm/s
9		Vibration velocity (z-axis)	mm/s
10	High speed shaft – Fixed bearing	Displacement (x-axis)	mm
11		Displacement (y-axis)	mm
12		Displacement (z-axis)	mm
13		Vibration velocity (x-axis)	mm/s
14		Vibration velocity (y-axis)	mm/s
15		Vibration velocity (z-axis)	mm/s
16		Temperature	°C
17	High speed shaft – Loose/float bearing	Displacement (x-axis)	mm
18		Displacement (y-axis)	mm
19		Vibration velocity (x-axis)	mm/s
20		Vibration velocity (y-axis)	mm/s
21		Temperature	°C

Table 3.2 - Available HS1000 low resolution operational parameters

#	Parameter	Units
22	Generator rotation speed	RPM
23	Output power	kW
24	Tidal flow rate (from EMEC model)	m/s

3.3.2 Data Analysis

Three different data analysis techniques were utilised within this study to analyse the data and identify suitable relationships and properties: correlation, principal component analysis and change point analysis.

The theory of each data analysis technique and results of each technique applied to the available data is explained in detail as follows.

3.3.2.1 Correlation

Correlation is a measure of dependency between separate data parameters. Within data mining, determining metrics of correlation in a dataset allows us to discover relationships between pairs of variables.

The covariance is a metric which describes how two variables change in relation to one another [50]. This can be calculated for two data parameters $\boldsymbol{x}$ and $\boldsymbol{y}$ as in (3.1), where n is the number of sampled data points, and $\bar{x}$ and $\bar{y}$ are the mean values of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. Large positive values of covariance between two parameters indicate an increasing linear relationship, and conversely, large negative values of covariance indicate a decreasing linear relationship between parameters.

$$cov(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{3.1}$$

However, the range of covariance values is dependent on the range of $\boldsymbol{x}$ and $\boldsymbol{y}$ values. Comparing covariance values for datasets containing many different types of measurement (from different sensors and components) is therefore of limited value. Covariance values can be standardized by dividing values by the standard deviations of $\boldsymbol{x}$ and $\boldsymbol{y}$. The resultant value is known as Pearson's correlation coefficient [51], denoted r_{xy} and calculated as in (3.2). Values of this coefficient range between +1 (describing an ideal increasing linear relationship) and -1 (describing an ideal decreasing linear relationship), where correlation coefficient values of zero describe two parameters having no linear association to each other.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.2}$$

A correlogram is a visualisation tool that displays the correlation coefficient between each individual parameter within a dataset [50]. This is a useful data mining tool to reveal dependencies when exploring large multivariate datasets. The correlogram is a $p\times p$ matrix, where p is the number of parameters in the dataset. Each row/column of this matrix corresponds to an individual data parameter. Values within the matrix represent the correlation coefficient between the parameters of each row/column. As $r_{xy} = r_{yx}$ the matrix is symmetric. Visualising the matrix as a two-dimensional grid, where correlation coefficient values are converted to a colour scale, allows the structure of a whole dataset to be visualised, clearly highlighting correlations and dependencies between individual parameters.

An example of a correlogram is shown in Figure 3.5 for a subset of HS1000 data parameters over a single downstream tidal cycle, including gearbox vibrations, generator rotor speed (RPM), output power and tidal flow rate. On the correlogram strong dependencies between individual parameters can be observed as bright and dark areas on the grid (representing strong increasing and decreasing linear relationships respectively). Levels of mid-grey represent little dependency between parameters. In this example, large positive correlations can be observed between vibration parameters, output power and tidal flow rate. This indicates these parameters are strongly related, e.g. as levels of output power and tidal flow rate increase/decrease, gearbox vibration increase/decrease accordingly. However, large negative correlations can be observed between rotation speed (RPM) and all other parameters. This indicates a strong inverse relationship between these parameters, prevalent during downstream tidal flows, when the turbine rotates in the opposite direction.

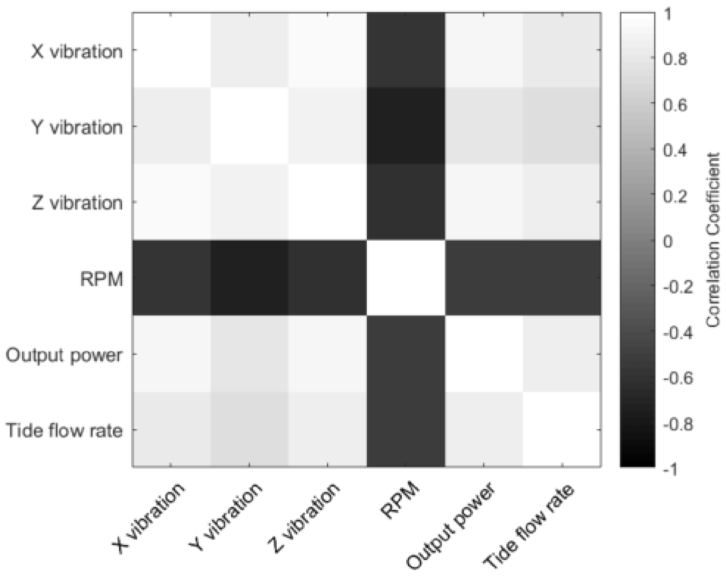


Figure 3.5 - Example of a correlogram for a subset of HS1000 data parameters over a single downstream tidal cycle

It is important to note that both covariance and Pearson's correlation coefficient measure the *linear* relationship between two variables. Certain *non-linear* relationships between variables that have clear dependency to one another therefore may not be revealed without further analysis.

3.3.2.2 Correlation Results

Pearson's correlation coefficient was calculated for every pair of data parameters, and visualised through a correlogram. Figure 3.6 and Figure 3.7 show correlograms for upstream and downstream tidal flow respectively, where parameters are labelled as in Table 3.1 and Table 3.2, p. 27.

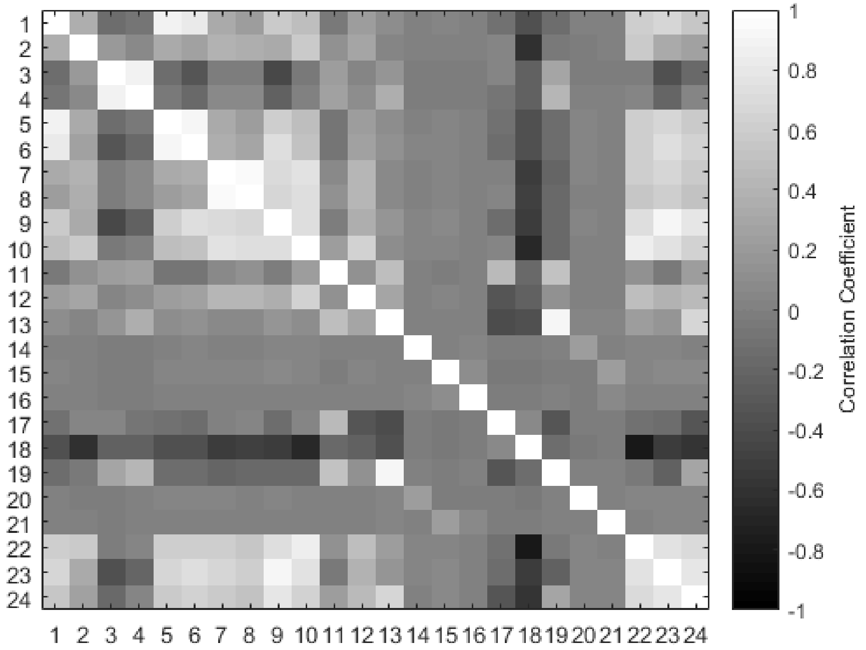


Figure 3.6 - Correlogram representing correlation between multiple condition parameters for upstream tidal flow

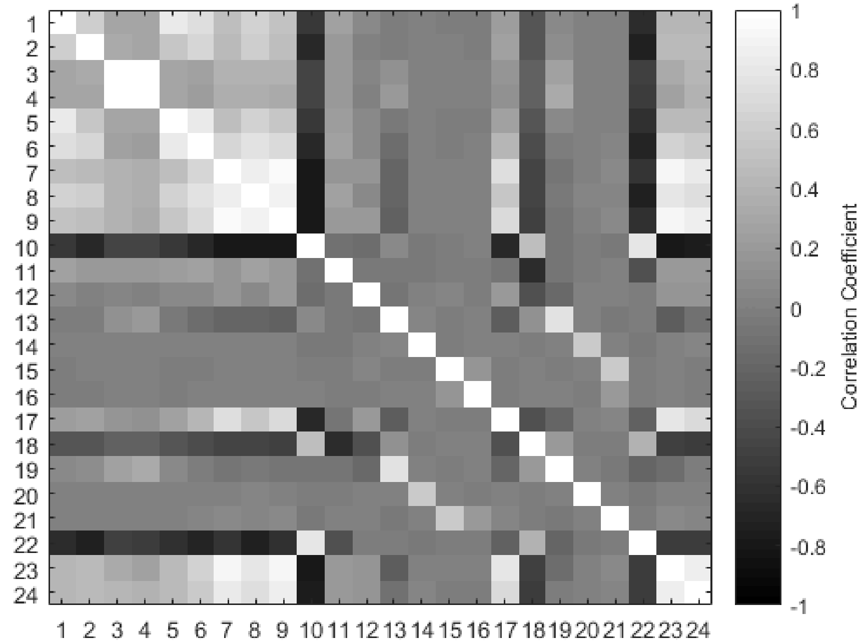


Figure 3.7 - Correlogram representing correlation between multiple condition parameters for downstream tidal flow

During upstream tidal flow, high positive correlations were observed between output power, generator rotation speed and tidal flow rate (parameters 22, 23 and 24 respectively). This was to be expected as the turbine increases its rotation speed and consequently output power in response to increasing tidal flow rates. High positive correlations were also observed between generator and gearbox vibration parameters (parameters 1 to 6 and 7 to 9 respectively). These vibration parameters were accordingly highly correlated to operational parameters output power, generator rotation speed and tidal flow rate (parameters 22, 23 and 24 respectively).

High speed shaft fixed bearing displacement (parameters 10 to 12, in particular x-axis parameter 10), and high speed shaft fixed bearing vibration x-axis (parameter 13), also have high positive correlations with gearbox and generator vibrations (parameters 1 to 9) and operational parameters (parameters 22, 23 and 24). Conversely, high speed shaft loose bearing displacement (parameters 17 and 18, particularly y-axis parameter 18) have negative correlations with generator and gearbox vibration parameters (parameters 1 to 9) and operation parameters (22, 23 and 24).

During downstream tidal flow the correlation between parameters are mostly similar, with high correlations between generator and gearbox vibrations, and operational parameters. However, some changes in correlation were observed. Firstly, generator rotation speed is highly negatively correlated with generator and gearbox vibrations (parameters 1 to 9) and operational parameters output power and tidal flow rate (parameters 23 and 24). This is to be expected as generator rotation speed is measured both positively and negatively, dependent on the direction of rotation.

Secondly, high speed shaft fixed bearing displacement x-axis (parameter 10) becomes highly negatively correlated to generator and gearbox vibrations (parameters 1 to 9), and operational parameters (parameters 23 and 24) during downstream tidal flow. In addition, high speed shaft loose bearing displacement x-axis (parameter 17) becomes positively correlated from negatively correlated to generator and gearbox vibrations (parameters 1 to 9), and operational parameters (parameters 23 and 24) during downstream tidal flow. These changes in bearing displacement correlations were due to axial forces on the high speed shaft acting in opposing directions as tidal flow changes direction from upstream to downstream.

As a result of these findings, modelling was focussed on firstly defining the relationships between generator, gearbox and bearing vibrations in relation to operational parameters. Bearing displacements were not modelled at this time due to differing responses between upstream and downstream tidal flows. This remains as future work.

3.3.2.3 Principal Component Analysis

Examining the relationships between pairs of variables allows us to find dependencies between sets of two parameters, however more complex relationships may be concealed further within the data.

Principal component analysis (PCA) [52] is a technique used to reveal these complex relationships by transforming data into projections along different axes. This is defined by weighted linear combinations of parameters, e.g. $2\boldsymbol{x}_1 + 3\boldsymbol{x}_2 + \boldsymbol{x}_3$ [50]. By performing this transformation, PCA removes linear correlations from data and generates a set of principal components.

Components are a list of coefficients, representing a weight for each input parameter, and an eigenvalue. Parameters with high weightings are the highest contributors to relationships within the data, and parameters with low weighting contribute the least [53]. A component's eigenvalue is representative of the how well a component describes the data as a whole.

Let $\boldsymbol{X}$ be an $n \times p$ matrix of p different data parameters $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_p\}$ each with n number of examples. The first principal component is required to have the largest possible variance in **X** [53]. Therefore, let $\boldsymbol{a}$ be a $p \times 1$ vector of the projection's weights resulting in the largest variance in **X**, where p is the number of parameters in the data. The variance of the projected data along $\boldsymbol{a}$ can be defined as in (3.3), where **V** is the covariance matrix of **X**.

$$\sigma_{\mathbf{a}}^2 = \boldsymbol{a}^T \mathbf{V} \boldsymbol{a} \tag{3.3}$$

The covariance matrix **V** of **X**, can be calculated as in (3.4), where **X** is mean-centred.

$$\mathbf{V} = \mathbf{X}^T \mathbf{X} \tag{3.4}$$

To get the first principal component, $\sigma_{\boldsymbol{a}}^2$ must be maximised. To do this, a normalisation constraint is set upon $\boldsymbol{a}$ vectors so that $\boldsymbol{a}^{\boldsymbol{T}} \boldsymbol{a} = 1$. With this constraint in place, the quantity to be maximised becomes the form as shown in (3.5), where λ is a Lagrange multiplier, a variable introduced to find the partial derivative.

$$u = \boldsymbol{a}^T \mathbf{V} \boldsymbol{a} - \lambda(\boldsymbol{a}^T \boldsymbol{a} - 1) \tag{3.5}$$

The maximum is found by differentiating u with respect to $\boldsymbol{a}$ as in (3.6).

$$\frac{\partial u}{\partial \boldsymbol{a}} = 2\mathbf{V}\boldsymbol{a} - 2\lambda\boldsymbol{a} = 0 \tag{3.6}$$

This expression can be reduced to the eigenvalue form as in (3.7) where **I** is the identity matrix. The first principal component is therefore the eigenvector associated with the largest eigenvalue in **V**.

$$(\mathbf{V} - \lambda\mathbf{I})\boldsymbol{a} = 0 \tag{3.7}$$

The second principal component is the orthogonal axis to the first component that has next the largest variance. This is therefore the eigenvector which corresponds to the second largest eigenvalue in $\mathbf{V}$ [50]. This is repeated for all subsequent principal components, where the k^{th} principal component is the eigenvector corresponding to the k^{th} largest eigenvalue in $\mathbf{V}$.

3.3.2.4 Principal Component Analysis Results

Principal component analysis was performed on the data following correlation analysis, allowing more complex structure within the data to be understood (i.e. beyond relationships between pairs of variables).

Table 3.3 and Table 3.4 show the results of principal component analysis for upstream and downstream tidal flows respectively, listing the weighting of each parameter in each principal component. Both upstream and downstream results are very similar.

In both cases, the first and second principal components are dominant, with eigenvalues of $1.3\text{x}10^5$ and $8.87\text{x}10^3$ respectively during upstream tidal flow, and $5.03\text{x}10^4$ and $1.51\text{x}10^4$ respectively during downstream tidal flow. These components are therefore most reflective of variance within the data. Within these components, generator rotor speed and output power (parameters 22 and 23) are weighted highly. This implies that generator rotor speed and output power are most reflective of the structure of the data as a whole, and drive the response of other parameters. Modelled tidal flow rate (parameter 24), however, is not identified as a leading operational parameter using PCA. This is primarily because flow rate is modelled and not directly measured. Therefore, the variance of this data parameter will not reflect the variance of other response parameters as well as measured output power or rotation speed.

Subsequent principal components have much lower eigenvalues, and therefore reflect less of the variance within the data. However, these components give strongest weightings to vibration parameters, firstly fixed and loose bearing x-axis vibration (parameters 13 and 19 respectively) followed by generator and gearbox vibrations (parameters 1 to 9).

These results mostly agree with the results from correlation analysis, where operational parameters were found to form relationships with vibration parameters from the generator, gearbox and bearings. However, PCA has excluded tidal flow rate as a leading operational parameter. As a result, modelling was focussed on defining vibration parameters in relation to the operational parameters of generator rotation speed and output power.

Table 3.3 - Parameter weights in each principal component for upstream data

(Columns correspond to principal components and rows corresponds to each parameter as in Table 3.1 and Table 3.2, p. 27. The final row gives the eigenvalues for each component.)

1	0.8	0.4	-1.7	-242.4	315.7	180.0	-262.4	-308.0	137.2	665.9	20.6	-157.1	0.2	-19.2	-342.9	-57.4	-112.6	127.3	-84.6	-20.1	-9.0	-16.9	-0.7	0.2
2	0.2	0.8	-5.3	20.5	24.1	22.9	-102.5	53.5	-27.9	205.4	-23.2	625.6	43.0	-50.5	-151.6	-159.9	419.5	-165.3	533.4	-61.2	36.7	66.0	9.1	0.4
3	-0.4	2.3	28.5	271.7	528.3	-247.1	152.3	55.7	-24.0	141.6	-26.2	488.5	41.6	29.0	228.4	51.0	-56.9	198.7	-445.2	11.2	7.7	16.1	16.6	-0.3
4	-0.2	1.8	58.3	281.1	560.2	-211.1	371.0	12.4	-1.2	-78.5	18.7	-416.2	-40.3	-28.6	-219.7	-49.5	50.9	-177.5	392.6	3.3	-5.8	-5.2	-12.4	0.4
5	0.8	0.3	-2.3	-271.7	362.5	219.5	-229.6	-88.5	29.5	-23.4	-8.5	-57.3	3.1	57.4	662.4	192.6	-17.0	-418.0	163.1	12.0	17.3	-9.3	-3.5	-0.4
6	1.4	-0.1	1.9	-412.5	372.6	478.3	59.6	302.3	-111.5	-442.4	-6.9	118.9	-3.2	-27.2	-279.5	-95.3	30.8	221.5	-82.3	-10.0	-14.3	4.0	2.6	0.1
7	1.1	0.3	-12.6	500.6	72.3	316.9	-384.1	-23.6	19.8	-94.4	23.5	-198.1	11.1	-23.5	-0.2	-117.7	512.1	-14.3	-246.7	326.2	-34.1	31.4	-0.6	-3.5
8	0.9	0.4	-9.4	531.7	-2.2	443.3	-161.7	64.2	-10.2	-15.7	-13.0	82.0	-11.8	28.8	-18.5	106.2	-466.8	56.0	191.7	-458.8	45.4	-8.4	8.3	3.1
9	2.2	-0.3	6.5	50.1	-170.3	521.1	697.8	39.0	-15.8	364.3	10.7	-3.0	18.4	8.9	147.6	7.1	109.6	-115.3	-86.6	142.4	-0.6	9.6	-4.0	-1.5
10	0.5	0.7	-6.0	80.4	-5.0	30.2	-79.5	96.6	-24.8	-1.5	1.7	198.7	-28.8	-34.0	-159.5	58.4	-412.8	-194.3	115.8	656.0	-177.0	-462.0	96.0	8.5
11	0.0	0.1	9.3	13.3	-3.6	5.2	-23.7	29.9	-7.2	9.5	-4.1	39.3	-6.2	0.8	-25.1	43.0	-165.5	-38.5	50.5	153.9	-664.0	660.4	-145.6	-201.8
12	0.1	0.2	4.6	18.0	-0.5	8.8	-36.2	27.5	-8.6	-17.2	6.2	43.6	-20.2	-19.9	-88.4	28.4	-232.9	-82.9	18.2	354.4	700.4	509.9	-216.6	37.3
13	1.3	4.2	909.8	-6.1	-40.3	48.5	16.7	-242.3	62.9	-128.4	-12.3	100.8	14.1	-14.5	-108.5	-30.1	-17.0	-209.9	-129.8	-63.9	-7.6	-4.6	4.0	3.7
14	0.0	0.0	0.7	-5.9	-6.1	-1.5	24.6	275.1	919.4	-33.1	-20.8	40.5	-270.6	-7.0	16.4	-33.5	13.2	-9.3	-8.8	-10.1	0.4	3.0	-0.7	0.5
15	0.0	0.0	0.1	-1.6	3.3	-2.5	5.0	-12.8	31.9	-20.2	852.2	28.1	71.0	-477.9	102.9	-148.2	-65.6	17.4	3.8	-27.6	-8.6	2.4	0.5	0.8
16	0.0	0.0	0.2	-1.8	-1.2	4.6	9.2	0.4	25.6	-8.3	162.0	46.8	-31.9	-79.1	-238.4	924.9	225.3	13.7	-11.5	-9.5	8.2	-1.8	-0.5	0.9
17	0.0	0.0	-1.7	2.0	-0.3	1.1	-2.6	3.2	-0.9	2.8	-1.5	5.2	1.0	2.7	1.9	5.7	-16.2	-0.6	7.5	13.5	-166.5	104.0	-83.5	976.7
18	0.0	-0.2	-1.6	-7.1	-2.2	-0.5	2.5	-3.1	1.4	-4.9	1.5	-24.5	-5.2	-0.3	-9.4	8.0	-35.5	-11.9	7.5	44.5	59.8	270.5	956.6	62.0
19	-0.6	2.3	396.2	-24.8	-23.5	-72.5	-159.3	663.7	-176.7	331.8	34.0	-211.4	-25.6	27.8	199.0	55.4	39.3	315.5	178.0	78.6	21.6	-11.5	2.9	1.1
20	0.0	0.0	1.7	-0.2	-0.1	-5.5	5.2	88.5	256.6	-27.4	-83.8	-52.9	953.9	-9.5	-18.7	44.8	-29.3	17.1	36.3	26.9	6.0	-2.0	1.0	-0.8
21	0.0	0.0	1.1	-0.9	2.1	-2.7	3.7	-8.8	30.9	-21.5	484.6	38.2	44.5	867.4	-71.3	-33.9	16.7	-11.3	22.2	26.4	3.0	-0.9	-0.6	-1.1
22	359.6	933.1	-5.0	-1.3	-2.3	-0.6	-0.3	-0.1	0.0	-0.6	0.0	-0.9	0.0	0.0	0.0	0.0	0.1	0.2	-0.2	-0.6	0.1	0.2	0.1	0.0
23	933.1	-359.6	0.8	0.4	0.5	-3.3	-0.6	1.2	-0.3	0.3	0.0	0.1	0.0	0.0	-0.1	0.0	-0.2	-0.1	0.0	-0.5	0.0	0.1	0.0	0.0
24	1.2	0.2	102.8	22.6	9.6	86.0	90.9	-446.8	126.5	-142.2	-30.6	39.0	-28.3	17.3	248.0	54.7	18.9	674.0	388.0	252.7	3.3	-14.7	-0.9	-2.4
Eig	**1.30E5**	**8.87E3**	**7.54**	**0.47**	**0.31**	**0.24**	**0.08**	**0.04**	**0.04**	**0.03**	**0.02**	**0.02**	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**2.7E-3**	**8.4E-4**	**1.2E-4**	**3.3E-5**

Table 3.4 - Parameter weights in each principal component for downstream data

(Columns correspond to principal components and rows corresponds to each parameter as in Table 3.1 and Table 3.2, p. 27. The final row gives the eigenvalues for each component.)

1	0.6	-0.6	-6.7	-3.0	296.0	-400.3	-363.3	49.1	34.7	-19.1	26.4	4.1	-131.5	272.7	-84.0	584.2	-141.9	329.7	-169.1	-37.5	-12.7	127.3	3.8	-9.4
2	0.6	-0.6	-12.5	3.3	46.9	-91.1	-281.8	105.9	-18.3	-327.9	-340.0	-1.3	615.4	-298.4	70.7	271.6	99.0	-293.6	83.9	-7.6	-3.3	-158.9	-4.2	18.7
3	1.5	-1.6	657.3	-182.5	7.5	-11.4	24.8	-65.8	8.6	59.9	-16.0	16.7	177.3	-261.5	80.1	-102.6	-179.0	484.3	-176.7	113.7	101.3	-295.7	6.4	-6.3
4	1.4	-2.0	713.9	-149.6	-12.9	8.9	-76.8	20.7	-8.5	-36.1	-8.3	-12.7	-146.9	239.3	-74.7	83.4	163.9	-447.3	173.6	-121.1	-98.3	294.7	-6.9	5.8
5	0.6	-0.5	-4.6	-6.2	303.6	-412.6	-371.5	49.6	32.7	39.4	136.8	-11.8	-328.6	-126.7	61.4	-387.4	86.5	-253.6	109.4	14.0	102.3	-450.8	3.3	5.0
6	0.6	-0.3	-26.4	-13.7	175.6	-233.1	-161.0	-18.1	15.6	21.7	19.0	19.2	109.1	-396.9	120.7	-369.3	-41.9	84.7	-103.2	26.1	-135.9	725.8	0.3	-21.1
7	1.7	0.7	23.3	-13.8	220.5	-196.1	484.4	-89.9	-36.7	-6.7	-211.1	-19.9	-222.3	-199.5	35.1	200.5	-36.8	-337.3	-464.5	344.6	-195.8	-61.3	21.8	-15.4
8	1.4	-0.3	-12.7	-41.2	296.2	-313.3	273.0	-164.6	-51.7	-4.0	-137.9	36.0	453.4	555.3	-186.2	-351.3	13.9	49.2	68.4	54.7	-63.1	-34.0	0.6	16.5
9	1.9	0.5	4.0	-34.8	250.5	-224.7	479.2	-63.0	-31.9	22.2	-7.0	9.1	-74.0	-284.1	79.7	220.0	100.6	106.9	463.5	-383.9	347.7	64.3	-21.8	10.5
10	-1.4	0.5	-10.8	-24.0	24.4	-60.5	19.4	-158.9	110.3	524.2	568.2	79.5	344.0	-124.0	4.3	212.6	77.5	-166.6	-12.8	-93.3	-334.9	-108.1	-55.5	69.5
11	0.1	-0.1	0.5	-25.8	-10.7	-8.1	-30.0	29.2	41.3	158.2	215.8	-4.1	133.4	45.0	-25.3	109.8	3.5	-140.0	129.3	639.8	603.6	172.8	225.3	29.5
12	0.0	0.1	0.4	-17.6	22.9	-14.4	23.3	77.4	24.0	-36.7	-82.4	-40.1	-127.9	-74.3	20.7	77.6	76.6	245.5	598.9	475.1	-541.9	-42.4	88.0	7.5
13	-0.2	-0.6	89.5	361.7	41.2	-9.8	44.4	536.4	341.7	537.9	-386.6	-68.5	42.9	28.9	-9.7	-41.1	-0.1	2.9	-8.9	-64.1	27.3	4.5	8.1	-0.9
14	0.0	0.0	0.6	8.1	529.0	465.3	-76.4	17.1	-7.2	15.1	-51.2	699.5	-39.8	-18.6	-47.2	4.0	2.9	-7.5	11.8	25.5	7.3	-3.5	2.0	-0.2
15	0.0	0.0	2.7	-9.9	0.0	4.9	32.5	-112.9	585.9	-218.4	53.0	2.4	-25.4	-42.9	-193.0	-10.3	694.5	186.0	-179.1	41.2	40.8	4.7	-2.4	2.7
16	0.0	0.0	1.8	1.9	8.2	-0.4	31.0	-66.4	217.0	-57.9	8.3	73.7	21.3	282.0	924.4	19.8	56.1	-1.1	-17.4	15.9	-4.3	5.0	3.8	-0.5
17	0.1	0.1	0.1	-2.8	0.0	4.1	1.0	20.0	-8.1	-36.8	-35.7	-5.6	-42.9	-3.2	5.2	-19.2	-17.2	23.6	-29.1	-5.4	1.2	20.0	-7.3	996.2
18	-0.1	0.0	0.0	10.0	-1.7	1.8	3.7	-35.7	-8.0	-7.7	-8.1	7.7	3.9	-13.5	0.6	-17.5	3.9	3.0	-60.5	-214.5	-99.2	-35.3	968.0	4.9
19	0.0	-1.6	214.6	898.9	10.0	-35.3	4.3	-255.0	-124.1	-186.7	155.6	24.4	14.0	-23.6	-0.5	8.8	1.3	22.5	25.4	49.6	-0.1	-1.8	-7.9	6.8
20	0.0	-0.1	-0.5	-1.3	542.9	443.4	-64.9	-77.8	-10.5	8.6	64.0	-699.3	47.9	10.8	45.3	3.3	7.1	3.6	-17.1	-15.5	-2.5	1.9	-1.2	-0.3
21	0.0	0.0	-2.1	-6.7	8.1	13.8	52.9	-156.4	673.0	-228.7	62.6	0.6	-6.1	-32.7	-134.2	-21.3	-625.4	-166.4	151.0	-41.0	-19.1	-1.9	-1.0	-0.4
22	-551.6	834.1	3.5	0.5	1.2	-1.3	-0.2	0.1	-0.2	-1.0	-0.9	-0.1	0.2	0.1	0.0	-0.2	0.0	0.0	0.2	0.1	0.6	0.3	0.1	-0.1
23	834.1	551.6	-0.2	1.2	-1.5	1.3	-2.2	-0.7	0.4	1.2	0.3	0.0	0.2	0.4	-0.1	-0.1	0.1	0.1	0.1	-0.1	-0.4	-0.1	0.0	-0.1
24	1.4	0.6	49.4	17.4	100.6	-30.2	246.1	710.8	-41.6	-392.7	485.7	7.1	96.3	33.0	11.1	-47.4	-37.2	17.4	-93.0	-7.0	-70.4	-13.0	10.1	-11.8
Eig	**5.03E4**	**1.51E4**	**0.95**	**0.32**	**0.07**	**0.07**	**0.05**	**0.04**	**0.03**	**0.03**	**0.02**	**0.02**	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**4.2E-3**	**3.2E-3**	**1.9E-4**	**1.4E-4**

3.3.2.5 Change Point Analysis

To form effective behavioural models, the key operational parameters within the data must be linked to the response of the turbine across a range of operating conditions. To make this connection, change point analysis was used, identifying points where multiple parameters changed together in time. These points were then used to define control modes of the turbine and clusters within the data. Here it is shown that this technique can be used as a method for identifying the control models of machinery directly from data, without prior knowledge of the system. This can be particularly useful in applications where the design details are unknown.

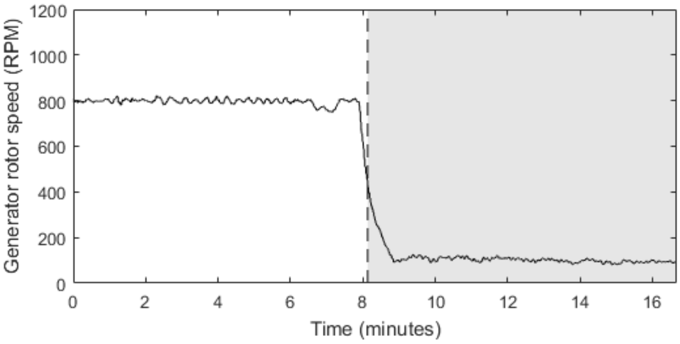
Change point analysis is a technique used to identify points in which data parameters change notably over time [54]. This can be used to define step changes in time series data or detect subtle changes in data missed by visual inspection. Analysing the change points of multiple different parameters allows different modes of operation to be found within the data. Separate parameters having similar change points are likely to be related to one another, linked to the same modes of operation within the turbine.

Change point analysis can be performed through calculating the cumulative sum of differences (CUSUM) of a data parameter from its mean value [54]. This is shown in (3.8) where C_i is the CUSUM at time step i, C_{i-1} is the CUSUM at the previous time step, x_i is the data value at time step i and $\bar{x}$ is the mean of the data. Changes in the gradient of C can be used to identify and define change points.

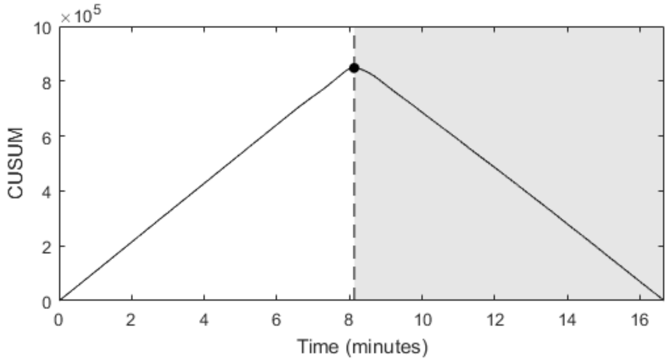
$$C_i = C_{i-1} + (x_i - \bar{x}) \tag{3.8}$$

Figure 3.8 and Figure 3.9 give examples of using CUSUM plots for detecting change points in HS1000 rotation speed data, indicative of changes in modes of the control scheme. Figure 3.8(a) shows the data for a short time window during turbine shutdown, with the corresponding CUSUM plot in Figure 3.8(b). A positive slope on a CUSUM plot indicates the parameter's values are above the mean value. Conversely, a negative slope on the CUSUM plot indicates parameter values are below the mean value. Here a maximum turning point on the CUSUM plot, where the gradient changes from positive to negative, detects the point in time where the turbine's mode of operation changes from generating to idle.

Figure 3.9 gives an example of using the CUSUM plot for detecting multiple change points in HS1000 rotation speed during a full tidal cycle, where generator rotation speed over time is shown in Figure 3.9(a) and the corresponding CUSUM plot is shown in Figure 3.9(b). Two turning points of the CUSUM plot (one maximum and one minimum) indicate change points indicating a change in control mode where the turbine rotation increases above 800 RPM and approaches its maximum rated value. Two additional change points can also be observed on the CUSUM plot (at 0.2 hours and 8.9 hours on the x-axis), where the gradient of C sharply changes but does not form a turning point. These points indicate where the turbine's operation changes from idle to operational, and vice versa.

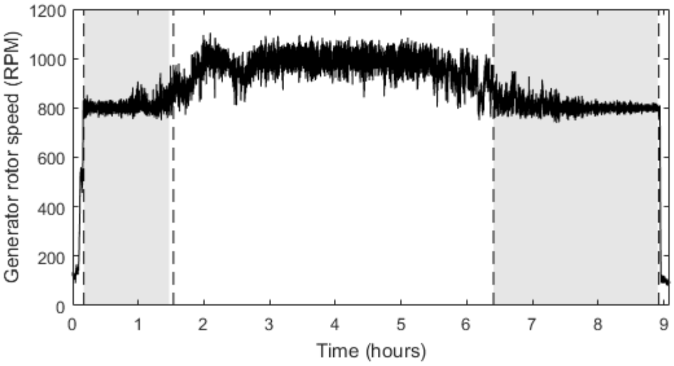


(a) Time-domain plot

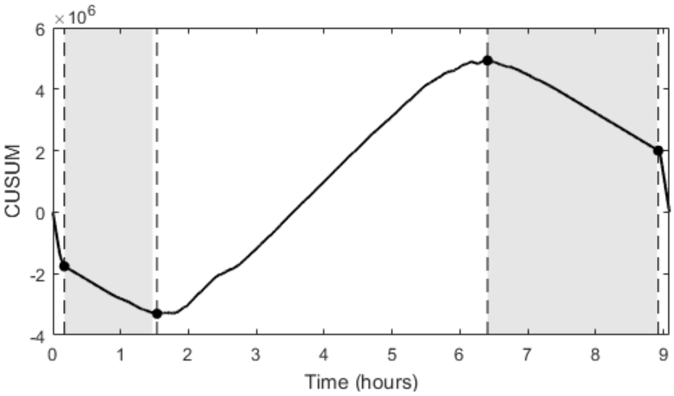


(b) CUSUM plot

Figure 3.8 - Change point analysis applied to short time window of generator rotation speed, where turbine operation changes from operational to idle



(a) Time-domain plot



(b) CUSUM plot

Figure 3.9 - Change point analysis applied to generator rotation speed over a full tidal cycle, indicating multiple changes in control mode

3.3.2.6 Change Point Analysis Results

Change point analysis was performed on the relationship between output power and generator rotor speed, in order to define the influence of the turbine's control scheme on these operational parameters.

Figure 3.10 shows how four control modes of operation are identified using CUSUM plots and change point analysis. Three change points are initially detected on the turbine's speed-power plot, as seen on the CUSUM plot in Figure 3.10(b). A fourth mode of operation was defined by iterating the change point analysis procedure to further split control mode 3, as in Figure 3.10(c).

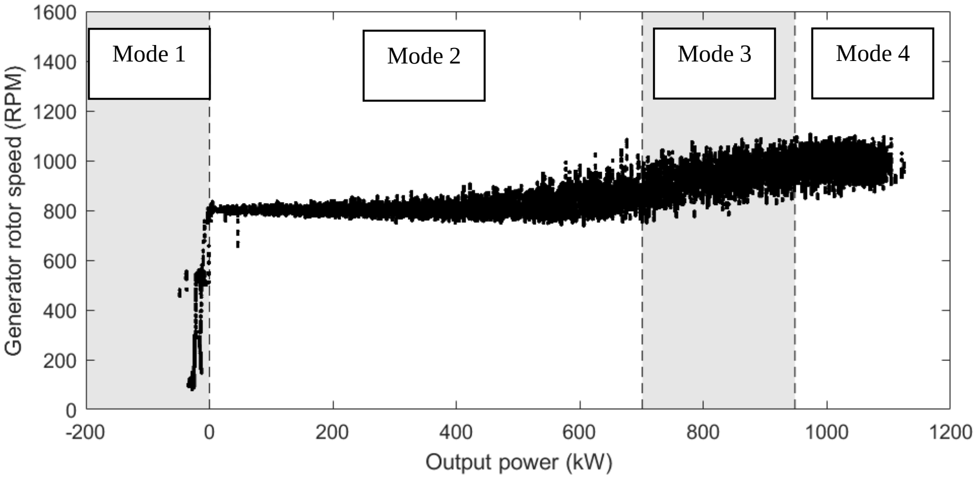
These modes occurred due to the following phases of turbine operation (confirmed through discussion with engineers):

1. Start up and shut down
2. Constant rotation speed as tidal flow increases (torque control)
3. Increase in rotation speed as tidal flow approaches peak velocity
4. Turbine rotor speed and power limitation (pitch and torque control)

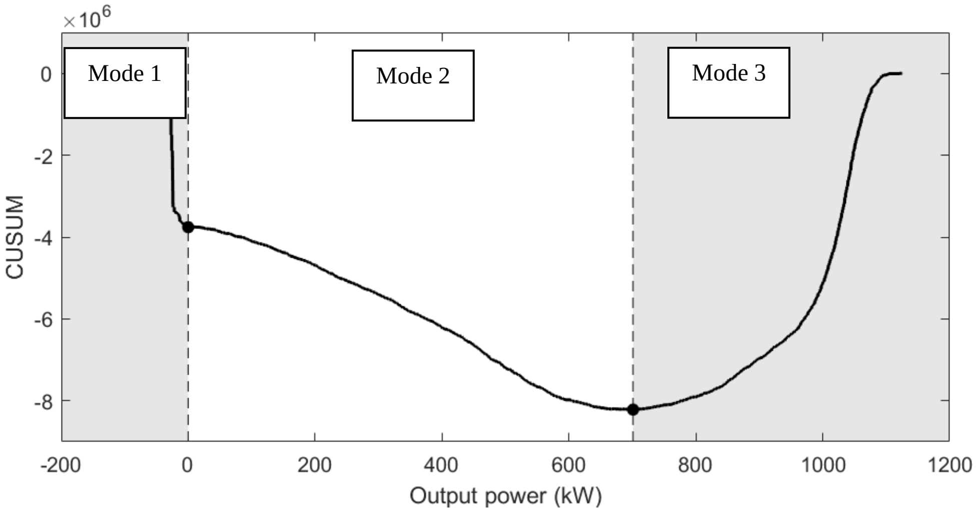
Separating these modes allowed the effects of the turbine's control scheme to be seen across other data parameters and was used to help partition data for use with anomaly detection techniques.

Figure 3.11 demonstrates how each operating mode shapes the trend of gearbox vibration against output power. Changes in operating state can be clearly seen as maximum and minimum turning points in vibration level. Maximum vibration occurs at the output side of the gearbox at ∼700 kW, where torque control is most active in the generator, holding the turbine's generator at 800 RPM as tidal flow increases. Beyond this point, vibration levels decrease as the effect of torque control decreases. Vibration then begins to increase as the turbine enters control mode 4, where the rotation speed of the machine is limited by pitch and torque control.

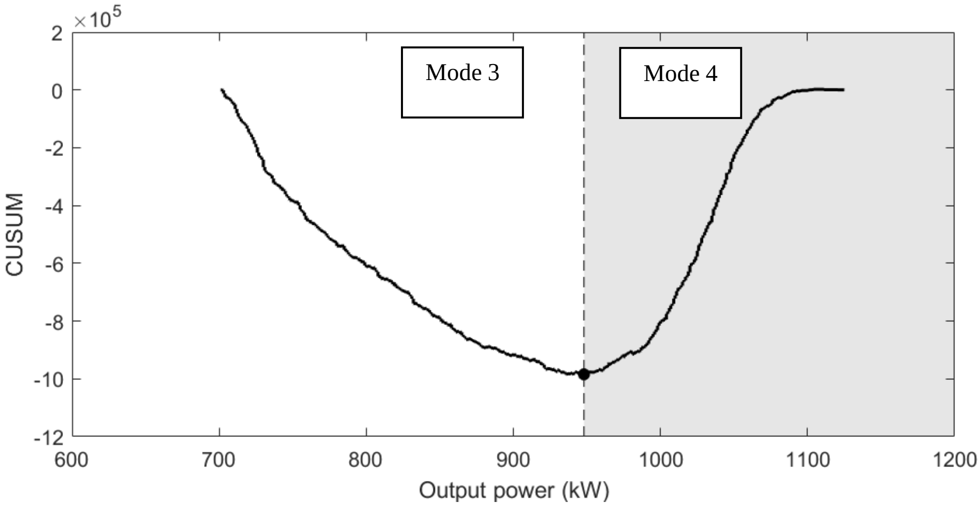
Figure 3.12 and Figure 3.13 show how overlapping clusters of data are formed in the trend of generator vibration against generator rotor speed during different phases of the turbine's control scheme. The general trend shows vibration increasing as generator rotation speed increases, however vibration levels are also shown to increase in control modes 2 and 4 (Figure 3.12), where pitch and torque control are most prevalent in controlling and limiting the turbine's rotation speed. These groups of data can be separated using calculated change points (Figure 3.13).



(a) Generator rotor speed against output power



(b) 1st iteration CUSUM plot



(c) 2nd iteration CUSUM plot

Figure 3.10 - Control modes defined on the turbine speed-power curve through point analysis

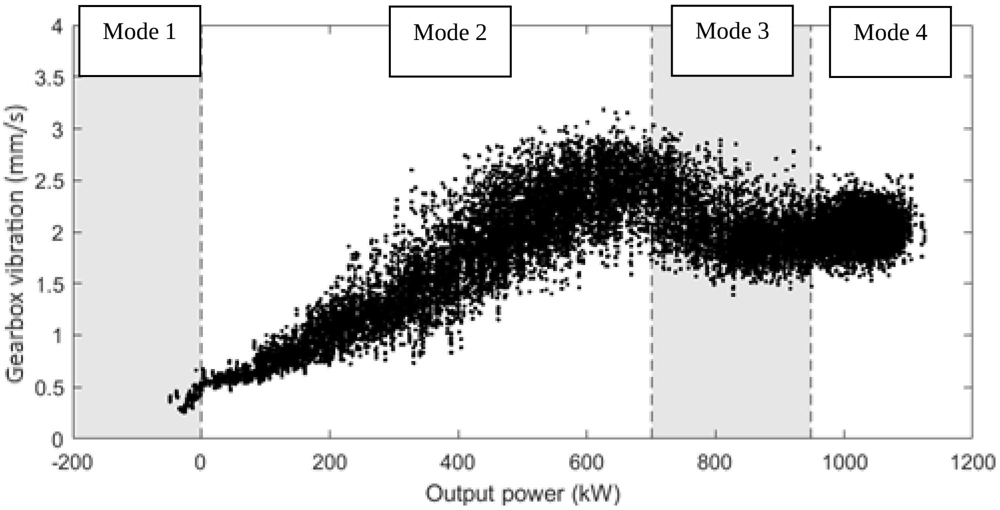


Figure 3.11 - Control modes on gearbox vibration (x-axis, output side) and output power relationship

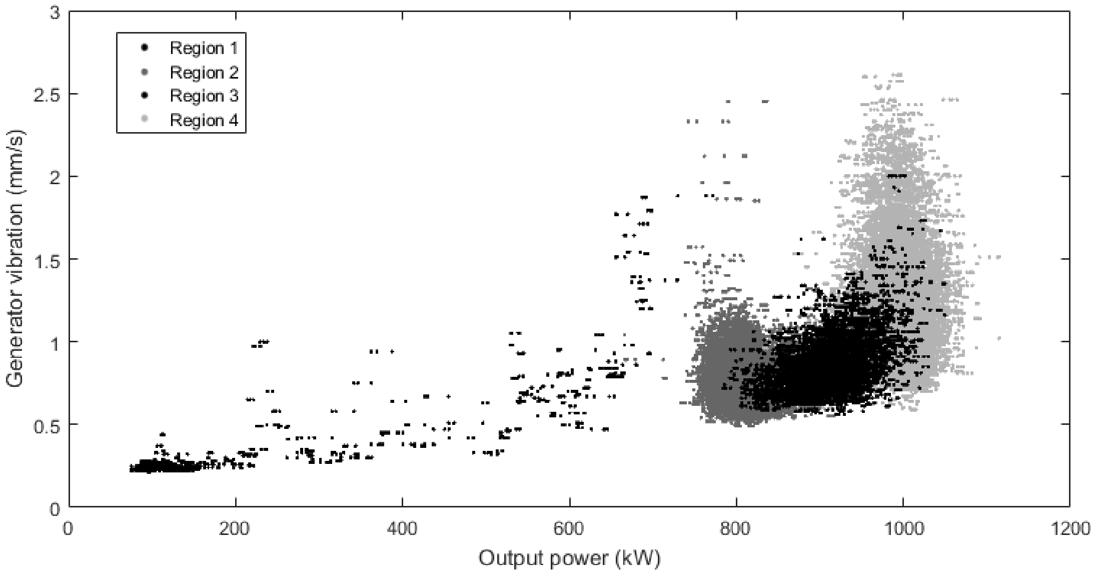


Figure 3.12 - Generator vibration (x-axis, input side) and generator rotor speed relationships control mode

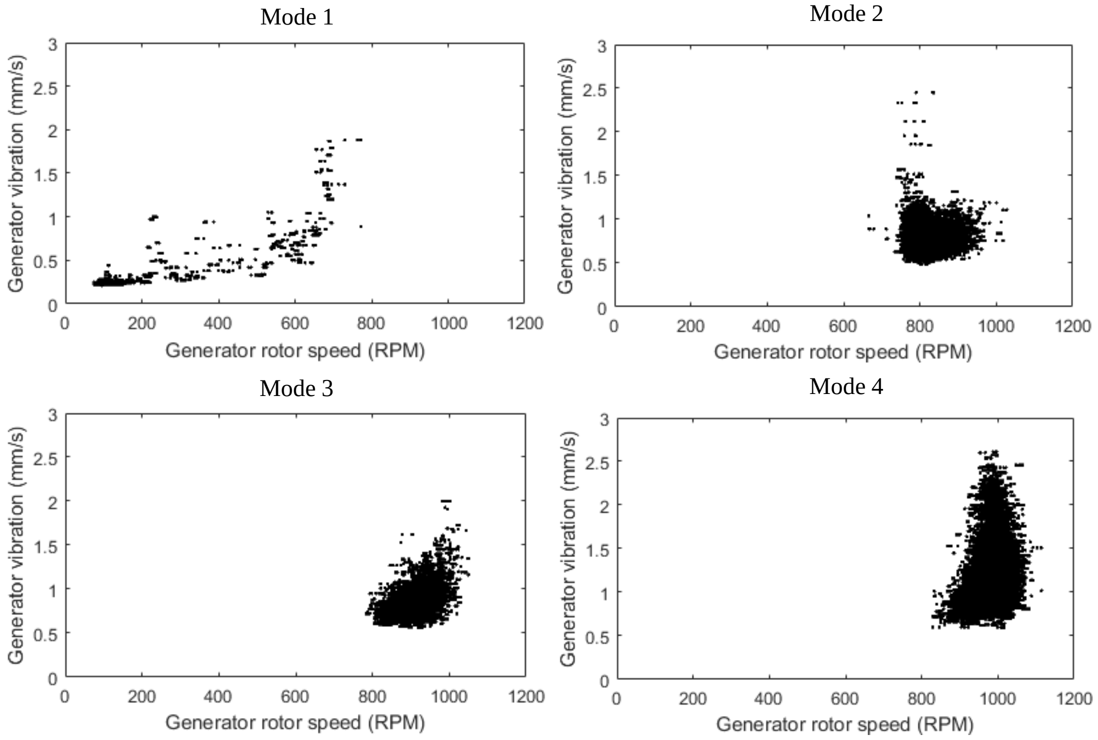


Figure 3.13 - Clusters of data from generator vibration (x-axis, input side) and generator rotor speed, separated by control mode change points

3.4 Data Preparation

The data preparation stage of the CRISP-DM methodology involves any processes to select, collate and/or organise data before the modelling stage [46].

The first step in the data preparation stage within this work was selecting and separating data parameters where relationships were revealed to be most indicative of the response of the turbine at the data understanding stage. This included vibration data parameters (measured from the turbine's gearbox, generator and high speed bearings) and generator rotation speed and output power. In addition, change point analysis revealed how different modes of the turbine's control scheme influenced these relationships, predominantly observable across generator and gearbox vibration.

The following relationships were therefore selected for the modelling stage:

- Gearbox vibration against output power
- Generator vibration against rotation speed (where data is clustered based on control mode)
- Bearing vibration against output power

Preparation steps here involved collating this data during tidal cycles in which the turbine was active and generating electricity for the grid. Data during periods in which the turbine was inactive over full tidal cycles were not collated for modelling. Data during periods where the turbine was partially active over a tidal cycle, e.g. shutdown events, were included for modelling.

In the case of generator vibration parameters, further preparation steps were included to label data clusters based on the control modes defined through change point analysis.

Finally, data was split into training and test datasets before modelling. 70% of the data was randomly sampled for training models, with the remaining 30% of data used for testing model performance in the evaluation stage. This is also known as hold-out cross-validation [55] and gives some protection against overfitting, where models fit closely to noise within the data, resulting in high performance against training data but poor performance against unseen data.

3.5 Modelling

The modelling stage of CRISP-DM involves selecting and applying various modelling techniques to best define the data based on the data mining goals [46].

In this work, the modelling stage was employed to find the modelling techniques that best describe the data within each relationship identified in the data understanding stage for anomaly detection.

3.5.1 Anomaly Detection Techniques

A number of different modelling methods can be used to perform anomaly detection. Pimentel, et al [56], have performed a comprehensive review of state of the art of anomaly detection (also known as novelty detection) techniques. Four categories of anomaly detection techniques were described as suitable for industrial monitoring and damage detection applications:

1. Probabilistic
2. Distance-based
3. Domain-based
4. Reconstruction-based

Techniques in each category are described as follows, identifying suitable techniques for modelling gearbox, generator and bearing vibration.

3.5.2 Probabilistic Techniques

Probabilistic techniques estimate the underlying probability density function (PDF) of data [56]. Thresholds are then used to define boundaries of normality within the data space, detecting anomalies within regions of low probability within a trained model. Parametric techniques describe a PDF through a set of variables, and include mixture modelling (predominantly Gaussian mixture modelling [57]) and state-space modelling (e.g. Hidden Markov Modelling [58]). Non-parametric techniques define a PDF through a model without set variables, such as kernel density estimation (KDE) [59]. Large datasets are required for reliable performance of these techniques, to ensure the underlying PDF is approximated accurately.

Within this work, probabilistic techniques were used to define clusters of generator vibration and rotation speed separated by the turbine's control mode, as in section 3.3.2.6, p. 38.

Parametric probabilistic anomaly detection methods fit a parametric model to data to estimate the underlying PDF. Within this method, the Gaussian distribution is commonly used as a model, however, other distributions such as the gamma, Poisson or Weibull are also used [56]. Complex distributions, where data does not precisely follow a pre-defined distribution, can be modelled through a mixture of distributions [60]. This is the principle of Gaussian mixture modelling (GMM), where multiple Gaussian distributions are combined to model the PDF of data [57].

D. P. Filev, et al. [61] show how GMMs can be used to define data clusters representing operating modes of a machine from vibration data for anomaly detection. This method was then used to automatically detect incipient and abrupt bearing faults where vibrations caused data to lie outside of operating mode clusters.

Non-parametric techniques, e.g. KDE, can be used in a similar fashion to parametric techniques, however, these techniques estimate the probability of data without a pre-defined parametric model [56]. These techniques have the potential to provide a better fit to data that have a more complex distribution that cannot be approximated accurately by a set function.

Q. Chen, et al [62], use kernel density estimation to identify anomalies in a glass melter machine and locate faults. A kernel density model is trained using principal component data from electrical power and temperature measurements. This paper shows KDE to out-perform parametric methods, showing that KDE is capable of detecting faults 50 minutes earlier than a normal density distribution.

Alternatively, state-space techniques, e.g. Hidden Markov Models, model the probabilities of transitions between different states [56] and not the probability of data points themselves. A. J. Brown, et al [63], use Hidden Markov Modelling to model the normal operating response of transformers, where anomalies are detected through deviations from this model. The model is trained using ultra-high frequency data describing partial discharge activity. This method does not directly model the data, and instead involves several additional processes to obtain the states for the HMM, obtained by clustering features from the high resolution data.

This technique, however, may not be suitable for modelling the transitions between clusters of data in this study, as data clusters separated by control mode have large overlapping regions (as seen in Figure 3.12, p. 41).

Instead, each cluster was modelled for anomaly detection independently using both parametric and non-parametric probabilistic techniques, namely GMM and KDE. The implementation of each technique is described in detail as follows. The fit of both GMM and KDE models were compared at the evaluation stage.

3.5.2.1 Gaussian Mixture Modelling (GMM)

Gaussian mixture models (GMMs) are parametric estimations of the probability density of data, where a complex distribution is modelled by a combination of individual Gaussian distributions [57].

This is commonly performed through the expectation-maximisation (EM) algorithm [64]. The EM algorithm involves two steps. First, the expectation (E) step calculates the probability of each individual Gaussian model in forming the distribution of measurements $\boldsymbol{x}$, given a current estimation of parameters $\boldsymbol{\theta}$. Secondly, the maximisation step (M) adjusts $\boldsymbol{\theta}$ to maximise the likelihood of $\boldsymbol{x}$ within the mixture model. This process is repeated iteratively until a convergence is reached.

A multivariate Gaussian distribution can be used to estimate the probability of $\boldsymbol{x}$ as in (3.9) where μ is the mean and $\boldsymbol{\Sigma}$ is the covariance matrix of the Gaussian model respectively. This can be expanded as in (3.10) to describe a mixture of K Gaussian models estimating the distribution of the same $\boldsymbol{x}$, where π_k is a weight for each Gaussian model.

$$f(\boldsymbol{x}|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.9}$$

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k f(\boldsymbol{x}|\mu_k, \boldsymbol{\Sigma}_k) \tag{3.10}$$

The parameters to be estimated in this problem are $\boldsymbol{\theta} = \{\pi_1, \dots, \pi_k, \mu_1, \dots, \mu_K, \sigma_1, \dots, \sigma_K\}$. Before starting the EM algorithm, these parameters are initialised. This can be done randomly, however more complex methods of parameter initialisation exist to improve the algorithm's performance [65].

Once $\boldsymbol{\theta}$ values are initialised, the probability that the kth Gaussian model is responsible for the distribution of $\boldsymbol{x}$ can be calculated as in (3.11). This is the E step of the EM algorithm.

$$\hat{P}(k|\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x}; \mu_k, \sigma_k)}{f(\boldsymbol{x})} \tag{3.11}$$

From here, the M step is performed, aiming to choose parameters which maximise a likelihood function of $\boldsymbol{x}$. The log likelihood of $\boldsymbol{x}$ given parameters $\boldsymbol{\theta}$ can be calculated as in (3.12).

$$l(\boldsymbol{x}|\boldsymbol{\theta}) = \ln p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln\left\{\sum_{k=1}^{K} \pi_k f(x_i|\mu_k, \boldsymbol{\Sigma}_k)\right\} \tag{3.12}$$

Parameters in $\boldsymbol{\theta}$ that maximise the log likelihood of data can be estimated as in (3.13), (3.14) and (3.15) over n data points in $\boldsymbol{x}$ [50].

$$\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n} \hat{P}(k|x_i) \tag{3.13}$$

$$\hat{u}_k = \frac{1}{n\hat{\pi}_k}\sum_{i=1}^{n} \hat{P}(k|x_i)x_i \tag{3.14}$$

$$\hat{\sigma}_k = \frac{1}{n\hat{\pi}_k}\sum_{i=1}^{n} \hat{P}(k|x_i)(x_i - \hat{u}_k)^2 \tag{3.15}$$

This is an iterative process, first using (3.11) to estimate $\hat{P}(k|\boldsymbol{x})$ from random $\boldsymbol{\theta}$ values, then updating $\boldsymbol{\theta}$ values using (3.13), (3.14) and (3.15). The process is stopped when some level of convergence is reached, i.e. the model parameters no longer change after each iteration.

Once an estimate of the data's PDF is achieved through a Gaussian model, an anomaly threshold can be set. This threshold value T is typically set as a value of probability, where if the probability of test point a is below the threshold value, as in (3.16), it is classified as an anomaly.

$$p(a) \leq T \tag{3.16}$$

Figure 3.14 shows a Gaussian mixture model defining the PDF of a data cluster of the HS1000's generator vibration and rotor speed in control mode 3, for $k = 5$. Here the mixture model has approximated the complex distribution of this data cluster through the combination of five individual Gaussians. Contour lines represent outwardly decreasing levels of probability, where data in the centre of the cluster is distributed more densely than data around the edge of the cluster.

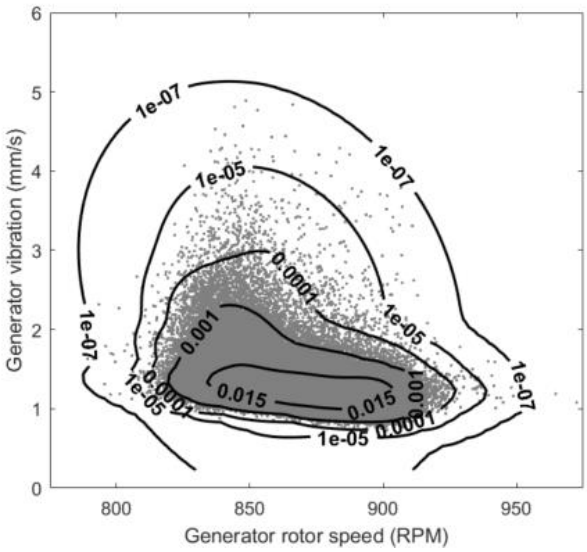


Figure 3.14 - Gaussian mixture model defining the PDF a data cluster between generator vibration and rotation speed, where contour lines indicate outwardly decreasing levels of probability

3.5.2.2 Kernel Density Estimation (KDE)

Kernel density estimation (KDE) [59] is a technique similar to Gaussian mixture modelling, where the PDF of data is estimated. However, this method is non-parametric and does not fit a pre-defined model (e.g. Gaussian) to define the PDF of the data as a whole. Instead, the underlying PDF of the data is estimated through the summation of kernel functions centred at each individual data point.

This process is expressed as in (3.17), where n is the size of data $\boldsymbol{x}$, K is a kernel function and h is a smoothing parameter known as the bandwidth.

$$\hat{P}_h(\boldsymbol{x}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - x_i}{h}\right) \tag{3.17}$$

The bandwidth parameter allows the kernel function to be scaled [66]. For example, when a Normal distribution is chosen as the kernel function, increasing the bandwidth parameter will decrease each kernel function's amplitude, but increase its variance. Choosing small values of h will provide a close fit to the data, however, choosing larger values of h spreads each kernel function and smooths the density estimate.

A common method of bandwidth selection is through the 'rule of thumb' method: a bandwidth that minimises the mean integrated squared error for Gaussian distributions [67]. This is expressed in (3.18) where $\hat{\sigma}$ is the standard deviation of the sampled data points.

$$h \approx 1.06\hat{\sigma}n^{-\frac{1}{5}} \tag{3.18}$$

As with GMMs, once a KDE model has been trained to estimate the data's PDF an anomaly threshold T can be set as a low value of probability, as in (3.16).

Figure 3.15 shows a KDE model defining the PDF of a data cluster between the HS1000's generator vibration and rotor speed in control mode 3. The bandwidth parameter h was automatically selected using the 'rule of thumb' method. In comparison with the GMM (Figure 3.14) the result is similar, however, the KDE model provides a tighter fit to data around lower density regions at the edges of the cluster. Where a probability threshold is set to define a boundary around the cluster, the KDE method may therefore allow for a more accurate anomaly detection model.

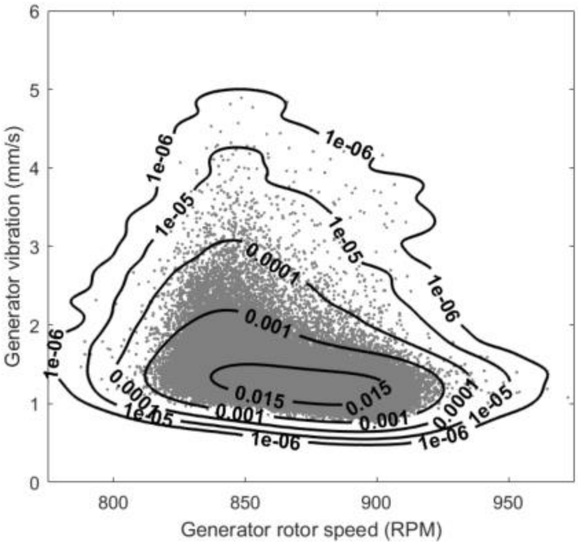


Figure 3.15 - KDE defining the PDF a data cluster between generator vibration and rotation speed, where contour lines indicate outwardly decreasing levels of probability

Both GMM and KDE models were tested in more detail at the evaluation stage, reported in section 3.6, comparing the GMM models with different k values and KDE models through the log likelihood (3.12) of data points from each model.

3.5.3 Distance-based Techniques

Distance-based techniques are equivalent to probabilistic PDF estimation, but use a distance metric rather than probability thresholds to identify anomalies within trained models [56]. Within this category, methods include clustering (such as k-means clustering [68]) and nearest neighbour (such as k-nearest neighbours [69]).

As with probabilistic anomaly detection techniques, distance-based techniques were most suitable for modelling data clusters in generator vibration at each control mode. The implementation of k-means clustering and k-nearest neighbour methods for modelling this data in comparison to probabilistic techniques is explained in detail as follows.

3.5.3.1 k-means Clustering

k-means clustering is a technique that assigns data points to one of a number of k different clusters, where k is pre-defined. Clusters are defined by their centroids, the clusters' central positions.

The k-means algorithm, also called Lloyd's algorithm [68] is an iterative process. Centroid values are first randomly initialised (although other initialisation methods exist [70]). The distances between each data point and each centroid value are then calculated. This distance can be calculated for each cluster using the Euclidean distance (3.19), where $\mathbf{X}$ is an $N\times p$ dimensional set of data points and $\boldsymbol{c}_k$ is a $1\times p$ vector describing the centroids of each dimension p in cluster k. Each data point in $\mathbf{X}$ can then be assigned to the cluster with the closest centroid.

$$\boldsymbol{d}(\mathbf{X},\boldsymbol{c}_k)=\sqrt{\sum_{i=1}^{p}(\boldsymbol{x}_i-\mathrm{c}_{ki})^2} \tag{3.19}$$

Once all data points have been assigned a cluster, the average of data points in each cluster in calculated. These averages are then assigned as the new centroid values for each k cluster. This process is repeated until the centroid values no longer change.

This technique can be used in unsupervised classification tasks, where k clusters are learned from unlabelled datasets. However, this technique can also be adapted for anomaly detection, where anomalies may be detected if they lie beyond a certain distance from the cluster centroids.

M. Lima, et al [71] use k-means clustering to detect anomalies within network traffic data. A baseline was generated using network traffic normal behaviour profiles, and anomalies were detected using the Euclidean distance between a sample and the cluster centroid.

Figure 3.16 shows k-means applied to a data cluster between the HS1000's generator vibration and rotor speed in control mode 3, where the centroid value of this cluster has been calculated using the Euclidean distance (3.19) and setting $k=1$ for a single cluster.

An anomaly threshold can be defined using this method as a distance value, yet, each dimension is measured in a different unit. A possible solution is to scale each different unit to a common range. This was achieved through Z-score standardisation [72], scaling each dimension to have zero mean and a standard deviation of 1. This is calculated for $\mathbf{X}$ as in (3.20), where $\boldsymbol{x}_p$ is an $N\times 1$ vector for each p dimension in $\mathbf{X}$, $\bar{x}_p$ is the mean of $\boldsymbol{x}_p$ and σ_p is the standard deviation of x_p.

$$\boldsymbol{Z_p}=\frac{\boldsymbol{x}_p-\bar{x}_p}{\sigma_p} \tag{3.20}$$

However, Figure 3.16 also shows data is not evenly distributed around the centroid in each dimension. Therefore, distance-based thresholds may not be as effective in comparison to probabilistic anomaly detection methods, where complex distributions can be modelled more accurately. In addition, this technique can be computationally expensive for large datasets [56], prevalent in this application where several months of operational data are used for training and testing models.

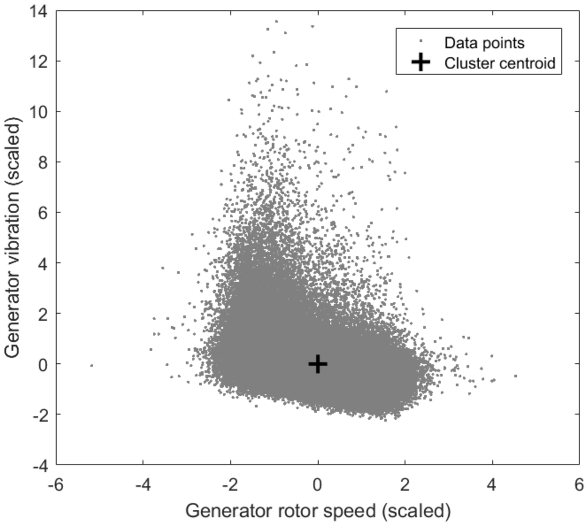


Figure 3.16 - k-means defining centroid of a data cluster between generator vibration and rotation speed, where anomalies are detected lying above certain distance from the centroid

3.5.3.2 k-nearest Neighbours

k-nearest neighbours (k-NN) is a non-parametric distance-based technique, where data points can be assigned to different clusters by selecting the majority cluster of the k training data points closest to the test data [69].

This technique can be used in anomaly detection, where an anomaly threshold can be applied to the mean distance between a data point and its k closest data points. As with k-means clustering, the Euclidean distance can be used as a distance metric. First, this distance is calculated in k-NN between a test data point $\boldsymbol{x}_{\text{test}}$ and all data points $\mathbf{X}$ in the model as in (3.21), where $\boldsymbol{x}_{\text{test}}$ is a $1 \times p$ dimensional vector and $\mathbf{X}$ is an $N \times p$ matrix. This can be used to identify the k data points closest to the test data point. From here, the mean distance of the k nearest data points can be calculated.

$$\boldsymbol{d}(\boldsymbol{x}_{\text{test}}, \mathbf{X}) = \sqrt{\sum_{i=1}^{p} \left(\boldsymbol{x}_{\text{test}_i} - \mathbf{X}_i\right)^2} \tag{3.21}$$

Figure 3.17 shows an example of this process for detecting anomalies compared to a data cluster between the HS1000’s generator vibration and rotor speed in control mode 3. An anomaly is detected if the mean Euclidean distance from its nearest k neighbouring data points (in this case $k = 4$) exceeds a pre-defined threshold. As with k-means, dimensions were scaled using Z-score standardisation [72], scaling each dimension to have zero mean and a standard deviation of 1, as in (3.20).

This technique is an improvement on k-means in this application, as data is not required to be evenly distributed. However, this technique also has its drawbacks. Firstly, as with k-means, it can be extremely computationally expensive for large datasets [56]. In addition, all data points within the model must be stored to apply the model to test data points for anomaly detection, as opposed to simply storing the centroid values as in k-means. More crucially for this application however, is that this technique does not generalise as well as probabilistic techniques and does not perform well where data is diversely distributed and arbitrarily shaped [56].

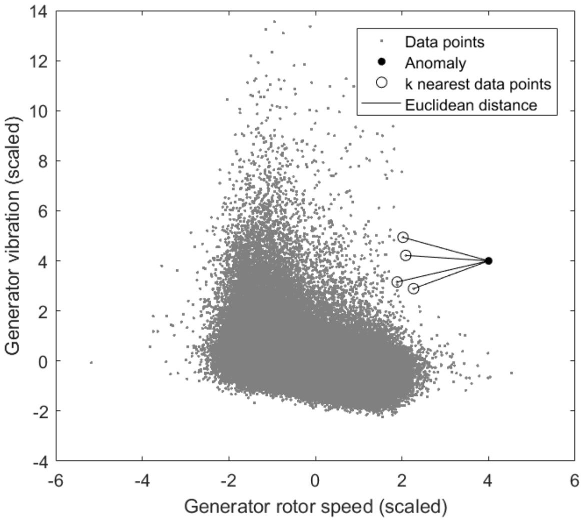


Figure 3.17 - Detecting an anomalous data point from a cluster of data points through k nearest neighbours using Euclidean distance with $k = 4$

Overall, probabilistic anomaly detection techniques are more suitable than distance-based techniques within this application in defining generator vibration data clusters. Probabilistic techniques can define complex distributions of data more effectively than k-means, and are less vulnerable to noise in the training data than k-NN. In addition, both techniques can be very computationally expensive, important in this case as training data describes the operation of the HS1000 over a period of several months. Distance-based techniques were therefore not considered further in the evaluation stage.

3.5.4 Reconstruction-based Techniques

Reconstruction-based techniques differ from probabilistic and distance-based approaches, detecting anomalies through models that aim to directly predict the values of data parameters [56]. This is also known as regression, where models estimate a function describing expected response of a system from a model trained on historical system data. A reconstruction error is then used to detect anomalies, describing the difference between test data points and the modelled system response. Reconstruction-based techniques include curve fitting [73] and neural networks (such as the multi-layer perceptron [74]).

In curve fitting, a model or function with parameters $\boldsymbol{\theta}$ is fitted to best approximate an output data parameter $\boldsymbol{y}$ from input data parameters $\boldsymbol{x}$ as $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\theta})$. This technique can be effective where data follows a clear function linked to the underlying response of the machine.

For example, R. Bi, et al. [75], use curve fitting to define the speed-power curve of a wind turbine's doubly-fed induction generator under normal conditions to detect faults from SCADA data. The results show a clearly interpretable model, where the curve function clearly expresses operating modes of the turbine. Faults causing anomalous behaviour can then be observed in comparison to this model and used to automatically trigger alarms.

Neural networks, however, aim to approximate output data from input data without a predefined parametric function. Instead, functions are learned through architectures connecting many neurons together, each applying a non-linear function to weighted combinations of input parameters [74]. This technique can be useful for defining complex non-linear responses that do not clearly follow a parametric function.

For example, M. Schlechtingen, et al [76], compare neural network approaches to simple linear curve fitting methods for anomaly detection of generator and gearbox faults within wind turbines. Linear curve fitting methods were found to be effective in defining the responses of data parameters that were clearly correlated (in this case generator coil temperatures), however, neural network approaches were found to be effective in defining more complex relationships where data weren't clearly correlated (in this case bearing temperatures).

Within this work, reconstruction-based anomaly detection techniques were most suitable in defining the response of gearbox vibration parameters again output power. Here, a clear relationship could be observed between data parameters, related to the underlying response of the turbine and its control scheme (identified in section 3.3.2.6, p. 38). Curve fitting approaches were therefore preferred in this case compared to neural network approaches, where the parameters of a function can be more easily interpreted and related to the dynamics of the system. The implementation of this technique through least squares regression to tune the parameters of a function to best fit data is described as follows.

3.5.4.1 Curve Fitting

Curve fitting involves fitting a pre-defined parametric model (or curve) to input data to estimate or predict output responses. This is described in (3.22), where f is a function of input data x_i at point i and model parameters $\boldsymbol{\theta}$, and $\hat{y}_i$ is the estimated response at point i.

$$\hat{y}_i = f(x_i, \boldsymbol{\theta}) \tag{3.22}$$

A technique used to solve this problem is least squares regression [77]. This technique is used to estimate the parameters $\boldsymbol{\theta}$ of a model to best fit observed data points, reducing the sum of squared errors of residuals. The residuals r_i (representing the prediction error) are the difference between observed responses y_i and predicted responses $\hat{y}_i$ for every i^{th} data point (3.23). The summed square error of residuals S is defined in (3.24) where n is the total number of observed data points.

$$r_i = y_i - \hat{y}_i = y_i - f(x_i, \boldsymbol{\theta}) \tag{3.23}$$

$$S = \sum_{i=1}^{n} r_i^2 \tag{3.24}$$

S is minimised by finding parameter values which set its gradient to zero. This is expressed as in (3.25) for models with sets of m different parameters in .

$$\frac{\partial S}{\partial \theta_j} = 2\sum_{j=1}^{m} r_i \frac{\partial r_i}{\partial \theta_j} = 2\sum_{j=1}^{m} r_i \frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j} = 0 \tag{3.25}$$

Linear models, consisting of a linear combination of parameters, can be expressed in matrix form (3.26), where $\boldsymbol{X}$ is an $n \times m$ matrix for the model and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of errors.

$$\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{\varepsilon} \tag{3.26}$$

From this notation, the least squares problem can be solved as in (3.27) to give an estimate of the model's parameters $\widehat{\boldsymbol{\theta}}$ [78].

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{3.27}$$

Conventional least squares regression methods assume errors are random and Normally distributed, with zero mean and a constant variance [77]. However, the method can be adapted to accommodate changes in variance across input parameters. This is known as weighted least squares regression and introduces scaling parameters (weights) w_i to the error estimate S as in (3.28).

$$S = \sum_{i=1}^{n} w_i (r_i^2) \tag{3.28}$$

Weight values control how much influence each observed data point y_i has on the fitting process, transforming the variance of responses to a constant value. A higher quality data point (with less variance) will therefore have more influence on the fit than lower quality data points (with higher variance). Weights can be calculated as in (3.29) for n data points where $\bar{y}$ is the mean of the output data parameter.

$$w_i = \left(\frac{1}{n}\sum_{k=1}^{n}(y_k - \bar{y})^2\right)^{-1} \tag{3.29}$$

Introducing weights to the error estimate changes the least squares solution in (3.27) to (3.30), where $\boldsymbol{w}$ is the diagonal elements of weight matrix $\boldsymbol{W}$.

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y} \tag{3.30}$$

The least squares method must be adapted for non-linear models, where such models can be given in the matrix form as in (3.31).

$$\boldsymbol{y} = f(\boldsymbol{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{3.31}$$

There is no closed-form solution to estimate parameters $\widehat{\boldsymbol{\theta}}$ to solve (3.25) for non-linear models. Non-linear least square algorithms therefore tune each parameter θ_j iteratively to minimize S, as in (3.32) where k is the current step and $\Delta\theta_j$ is the step size. Examples of such techniques include the trust-region [79] and Levenberg-Marquardt [80] algorithms.

$$\theta_j^{k+1} = \theta_j^k + \Delta\theta_j \tag{3.32}$$

3.5.5 Domain-based Techniques

Domain-based techniques create an anomaly boundary from the structure of training data [56]. The location of this boundary is determined using only data closest to the boundary, and does not rely on the properties of the data distribution as a whole.

This category of anomaly detection technique was used in this work to define the response of high speed bearing vibrations against output power. The relationships between these parameters could be observed to follow a trend only across maximum levels of vibration, where peaks occurred at different values of output power. Domain-based anomaly detection techniques were therefore most suitable in this case in comparison to probabilistic, distance-based or reconstruction based, as only the data across maximum levels of vibration form this trend.

Techniques within this category include support vector machines (SVM), where anomaly boundaries are set to define the structure of data based on parametric functions [81]. However, an alternative technique, referred to in this work as ‘envelope fitting’, was used to define the response of bearing vibrations against output power. These methods are described in more detail as follows.

3.5.5.1 One Class Support Vector Machines (SVMs)

Support vector machines (SVMs) are commonly used as a data classification method, where a hyperplane is defined to separate data classes in multidimensional space from a labelled dataset [82]. This hyperplane is defined to maximize the distance between the hyperplane and the nearest data point (the margin), giving optimum separation between classes, Figure 3.18. The support vectors are the data points closest to the hyperplane, and define the boundaries between classes. The hyperplane of SVMs is natively linear, but can be made non-linear through the ‘kernel trick’: representing input data in a higher dimensional space through a kernel function [83]. Classification SVMs are described in more detail in chapter 4.

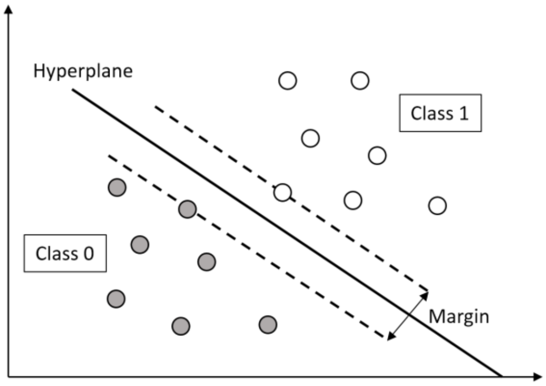


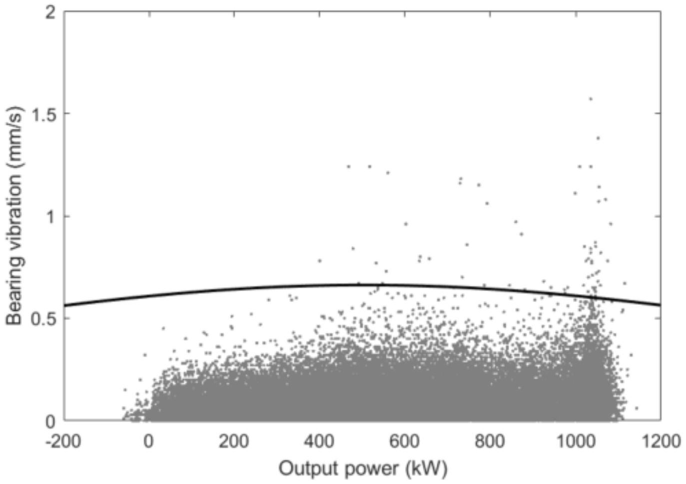
Figure 3.18 - A linear support vector machine for binary classification

Classification SVMs can be adapted for anomaly detection through *one-class* SVMs [81]. This adaptation defines a novelty boundary corresponding to a kernel to separate data from the origin of the high dimensional kernel space. In this method, a fixed percentage of the data is required to be defined as anomalous before training the classifier.

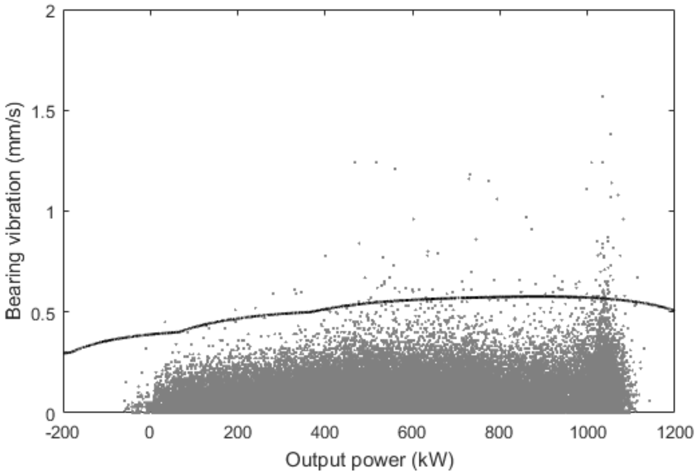
A. Purarjomandlangrudi, et al [84], use an SVM to detect anomalies within bearings of a wind turbine. Kurtosis and non-Gaussianity score parameters detailing statistical properties of vibration data were recorded from a bearing test rig and used to train a one-class SVM. Anomalies from various bearing faults, including inner race, outer race and ball faults, were distinguished from normal behaviour using this technique.

Figure 3.19 (a) and (b) shows a one-class SVM trained to fit an anomaly boundary on bearing vibration data against output power for polynomial and kernel functions respectively, with 0.1% of the data defined as anomalous.

It can be seen that although this technique defines a region separating the majority of data points from a fraction of data points lying furthest from the centre, it does not fit closely to levels of vibration varying over the range of output power, particularly the peak of vibration about 1000 kW. Changing the percentage of anomalous data points had little effect on changing the shape of each boundary for this data.



(a) Polynomial kernel function



(a) Gaussian kernel function

Figure 3.19 - One-class SVM with different kernel functions defining anomaly boundary for fix bearing vibration (X-axis) against output power

3.5.5.2 Envelope Fitting

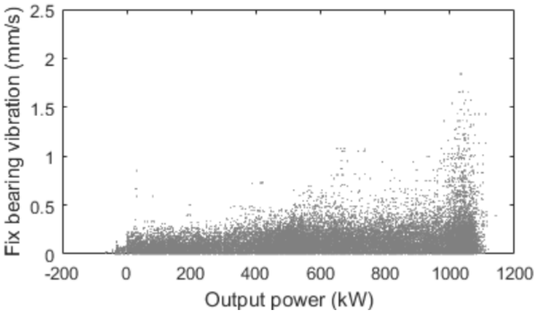
Due to the limitations of SVM domain-based anomaly detection, an alternative approach was taken to define an anomaly boundary for high speed bearing vibration. Within this work, this approach is referred to as ‘envelope fitting’, where an anomaly boundary was defined by fitting a parametric curve over the maximum values of bearing vibration against output power.

Firstly, vibration parameters are sorted by increasing output power. The envelope of the vibration response is represented through $(\boldsymbol{x}_{\mathrm{env}}, \boldsymbol{y}_{\mathrm{env}})$ by sampling maximum values of vibration at discrete steps of output power. The step size was set to the highest resolution of output power measurement, in this case 1 kW. However, this step size could be increased (for example to 5, 10, 50 kW, etc.) resulting in a smoother, but less accurate, approximation of the maximum vibration envelope.

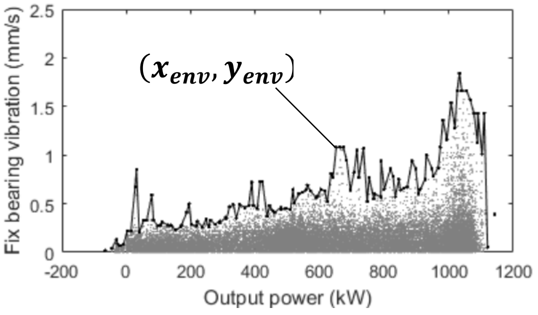
Finally, curve fitting was then applied to the sampled envelope data points, fitting a parametric model $f(\boldsymbol{x}_{\mathrm{env}}, \boldsymbol{\theta})$ to define the maximum boundary of vibration across the profile. This is implemented using least squares regression (as in section 3.5.4.1, p. 52).

Figure 3.20 describes this process for fix bearing vibration (X-axis) trended against output power. By concentrating on only defining the maximum levels of vibration, this method provides a much closer fit to data in contrast with one-class SVM models (Figure 3.19).

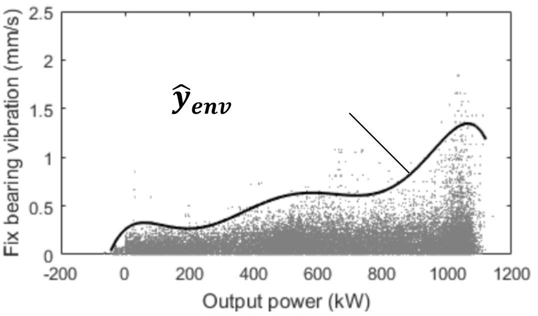
Anomaly detection is then performed by using a trained envelope model for each vibration parameter, where points lying above the anomaly boundary, or beyond a threshold set at distances in relation to the anomaly boundary, are considered anomalous.



(a) Step 1: Vibration parameter sorted by increasing output power.



(b) Step 2: Maximum vibration values sampled at each step of output power.



(c) Step 3: n order function fitted across maximum values

Figure 3.20 - Example of the envelope fitting process for fix bearing vibration (x-axis)

3.6 Evaluation

The evaluation stage of CRISP-DM involves reviewing how models achieved through the data mining process comply with the objectives set in the business understanding stage [46]. At a low level, this may involve evaluating the performance of each model in completing the business objectives. However, at a higher level, this may involve reviewing the full data mining process and redefining the business or data mining objectives should this process reveal important business issues that were not originally considered [46].

In this work, the evaluation stage involved a low level assessment of the performance of each model in defining the normal response of the turbine over each relationship. This involved using metrics to measure the fit of each model to test data and changing parameters of each model where necessary to improve each model's performance. Finally, anomaly thresholds were applied to each model before deployment.

Firstly, curve fitting was used to model the relationship between gearbox vibration parameters and output power. Different functions were trained and tested to best define this relationship, with anomaly thresholds set based on the models' reconstruction errors.

Secondly, the clusters of data between generator vibrations and generator rotor speed (revealed through change point analysis) were modelled using probability-based anomaly detection techniques. Both Gaussian mixture modelling (GMM) and kernel density estimation (KDE) were compared, where KDE was found to best model the data. Anomaly thresholds were then set based on a level of probability describing normal operation.

Finally, envelope fitting (a domain-based technique) was used to define the response of bearing vibration parameters to output power. A number of functions were evaluated to best fit these envelopes and define a boundary of normal behaviour. Anomaly thresholds were set based on these boundaries.

3.6.1 Gearbox Vibration

Data analysis revealed a relationship between gearbox vibration and output power, where all data points followed an underlying function, with local maximum and minimum turning points reflective of different modes of the control scheme (section 3.3.2.6, Figure 3.11, p. 40).

Maximum vibration occurs in control mode 2, where the difference between torque at the turbine's input and generator is at its highest, as turbine rotation speed is held fixed at 800 RPM as tidal flow rate increases. Beyond this point, gearbox vibration decreases as the different in torque decreases, where the control scheme allows generator rotation speed to increase with tidal flow rate. Vibration then begins to increase once more when the turbine's rotation speed is further limited at 1000 RPM at high tidal flow rates.

Curve fitting was used as the modelling technique in this case. This is a reconstruction-based anomaly detection technique, where a parametric model was fitted to data $(\boldsymbol{x}, \boldsymbol{y})$ to model a function $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\theta})$. This was performed through weighted least squares regression (section 3.5.4.1, p. 52), where parameters $\boldsymbol{\theta}$ of a model were tuned to best fit the data.

A number of parametric functions were first trained and tested to best define the data, including Gaussian and polynomial functions of varying orders, and a piecewise linear model. Once functions that best fit the data were found for each vibration parameter, anomaly thresholds were applied to the prediction error from test data, where bounds of acceptable deviation from each model were set from estimations of the prediction errors' probability density function (PDF).

3.6.1.1 Model Selection

Models were first trained to best fit the normal vibration response of the gearbox from historical data from the HS1000, from a period of operation between September 2013 and February 2014. 70% of the data were randomly sampled for training, with the remaining 30% used for testing. Separate models were used to define vibration over the X, Y and Z axes.

A number of different types of functions were tested including Gaussian and polynomial functions of different orders, as well as a piecewise linear fit within each control scheme mode as defined in section 3.3.2.6, p. 38. Some examples are shown in Figure 3.21. An n^{th} order Gaussian function is shown in (3.33), where the parameters of the model are $\boldsymbol{\theta} = \{a_1, \dots, a_n, b_1, \dots b_n, c_1, \dots c_n\}$. An n^{th} order polynomial function is shown in (3.34), where the parameters of the model are $\boldsymbol{\theta} = \{a_0, \dots, a_n\}$. A linear model is given by (3.34) where $n = 1$.

$$y = \sum_{i=1}^{n} \left(a_i e^{-\left(a - \frac{b_i}{c_i}\right)^2} \right) \tag{3.33}$$

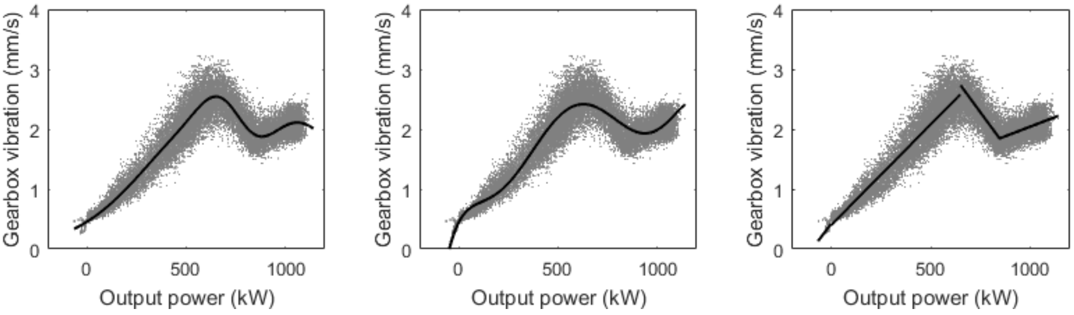
$$y = \sum_{i=0}^{n} a_i x^i \tag{3.34}$$

The fit of each function to test data was compared using the root mean squared error (RMSE) and R^2 values. The RMSE is calculated for n data points as in (3.35), where y_i is the measured response and $\hat{y}_i$ is the modelled response at point i. The R^2 value is calculated as in (3.36), where $\bar{y}$ is the mean measured response.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.35}$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.36}$$

Table 3.5 shows the values of these metrics calculated for the test data and each model. In each case Gaussian models were found to best fit the data, where a 3rd order Gaussian was best reflective of X-axis vibration and 6th order Gaussians were best reflective of both Y-axis and Z-axis gearbox vibration.



(a) 3rd order Gaussian (b) 6th order polynomial (c) Piecewise linear

Figure 3.21 - Curve fitting using different functions

Table 3.5 - Goodness of fit metrics for gearbox vibration model test errors

Function	Order	X-axis		Y-axis		Z-axis	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
Gaussian	3rd	**0.242**	**0.878**	0.263	0.849	0.241	0.908
	4th	0.248	0.872	0.255	0.858	0.242	0.907
	5th	0.243	**0.878**	0.263	0.849	0.242	0.907
	6th	0.243	0.877	**0.251**	**0.862**	**0.237**	**0.911**
Polynomial	3rd	0.262	0.857	0.277	0.832	0.287	0.869
	4th	0.261	0.858	0.276	0.834	0.273	0.882
	5th	0.260	0.859	0.276	0.834	0.271	0.883
	6th	0.260	0.859	0.267	0.844	0.248	0.902
Piecewise Linear	----	0.250	0.870	0.265	0.845	0.260	0.892

3.6.1.2 Anomaly Thresholds

The distribution of the prediction error under healthy turbine conditions was used to set fault thresholds. Prediction errors exceeding these thresholds would indicate a meaningful change in the vibration response of the turbine's gearbox.

Kernel density estimation was used to estimate the distribution of prediction errors, r_i calculated as in (3.23), p. 52, generated from healthy turbine data. This process estimates the underlying PDF of data from limited observed points. An estimated PDF is calculated through the summation of kernel functions at each observed data point [59].

Thresholds were applied to each kernel density estimate based on the work described in [85]. A lower limit, $L(k)$, and upper limit, $U(k)$, are applied to each estimated distribution, k, as in (3.37) and (3.38), where $mo(k)$ is the mode of the distribution, $\sigma(k)$ is the standard deviation of the distribution and A is a tunable domain-specific free parameter.

$$L(k) = mo(k) - A\sigma(k) \tag{3.37}$$

$$U(k) = mo(k) + A\sigma(k) \tag{3.38}$$

Faults within the gearbox resulting in an increase in measured vibration will increase the prediction error from the model. The A parameter is therefore chosen to set limits which describe acceptable deviation from each model, beyond which a fault is likely to have occurred.

In order to reduce alarm responses due to single point vibration anomalies, prediction errors r_i were averaged per minute using a moving average filter (3.39), where N is the filter window size, before being compared to the thresholds.

$$m_i = \frac{1}{N} \sum_{k=i-\left(\frac{N-1}{2}\right)}^{i+\left(\frac{N+1}{2}\right)} r_k \tag{3.39}$$

Figure 3.22 shows the kernel density estimate for the prediction errors for the X-axis gearbox vibration model, where anomaly thresholds have been set as in (3.37) and (3.38). Similarly, threshold values were set for Y-axis and Z-axis models. Table 3.6 summarises these results.

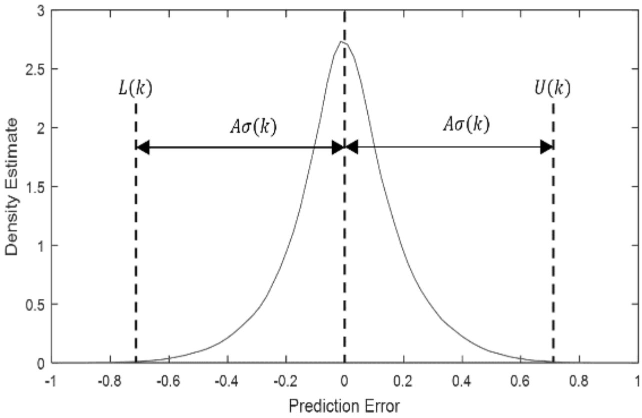


Figure 3.22 - Thresholds of expected deviation applied to kernel density estimation of model prediction error

Table 3.6 - Anomaly thresholds for gearbox vibration model prediction errors

Model	$\boldsymbol{L(k)}$ **(mm/s)**	$\boldsymbol{U(k)}$ **(mm/s)**
X-axis	-0.75	0.75
Y-axis	-0.80	0.80
Z-axis	-0.78	0.78

3.6.2 Generator Vibration

Data analysis revealed a relationship between generator vibration parameters and operational parameters generator rotor speed and output power. Through defining phases of the turbine's control scheme, clusters of data were identified and separated in the relationship between generator vibration and generator rotor speed (section 3.3.2.6, Figure 3.12 and Figure 3.13, p. 41).

Probabilistic-based anomaly detection techniques (section 3.5.2, p. 43) were used to model these clusters of data driven by the turbine's control scheme, where boundaries around groups of data points can be defined through an estimation of the data's underlying probability density function (PDF).

Two probability-based anomaly detection techniques were compared at the evaluation stage: Gaussian mixture modelling (GMM) and kernel density estimation (KDE), finding KDE to be most suitable. Anomaly thresholds were then applied to KDE models using each density estimate's 99% confidence interval.

3.6.2.1 Gaussian Mixture Modelling and Kernel Density Estimation

Two probability-based anomaly detection techniques were tested to define clusters of generator vibration and generator rotor speed data, formed at each mode of the turbine's operating scheme.

Firstly, Gaussian mixture modelling was tested. This technique approximates the PDF of data by fitting K multi-dimensional Gaussian functions to data. This was performed through the Expectation-Maximisation algorithm, as described in section 3.5.2.1, p. 44.

Secondly, kernel density estimation was tested. This technique aims to more accurately estimate the PDF of data through the summation of individual kernel functions at each data point, described in section 3.5.2.2, p. 46, adapting an implementation by Botev, et al. [86].

The effectiveness of both techniques were evaluated through measurement of the log likelihood for test data in each model, calculated as in (3.12), p. 45, where higher values of log likelihood indicated a closer fit to the data. Table 3.7 shows these results, for GMM models with differing k values and for KDE models.

These results show that KDE returns the highest log likelihood values for each parameter and control mode. This indicates that KDE models provide a closer fit to the data in each cluster as data points are modelled with higher probability. This can also be observed in Figure 3.23 and Figure 3.24, where the contours of KDE models fit closer around data clusters in comparison to GMM models. In addition, the outermost contours of probability are lower in GMM models compared to KDE models., showing data points have higher levels of probability in these models.

As a result, KDE was chosen to model the response of generator vibration against generator rotor speed in each mode of the turbine's control scheme.

Table 3.7 - Log likelihood for GMM and KDE generator vibration models

Generator side	**Vibration axis**	**Control mode**	**GMM**					**KDE**
			$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	
Input	X-axis	1	3.31	6.05	6.11	6.16	6.20	**6.52**
		2	10.8	11.2	11.2	11.3	11.3	**11.5**
		3	8.23	8.27	8.3	8.3	8.3	**9.17**
		4	7.1	7.3	7.36	7.37	7.37	**7.42**
	Y-axis	1	2.93	4.18	5.46	5.48	5.77	**5.82**
		2	10.4	**10.7**	**10.7**	**10.7**	**10.7**	**10.7**
		3	8.47	8.59	8.59	8.6	8.6	**8.93**
		4	7.44	7.69	7.75	7.78	7.79	**7.81**
	Z-axis	1	2.51	4.58	4.73	4.74	4.85	**5.00**
		2	9.37	9.71	9.73	9.73	9.75	**9.80**
		3	7.86	8.29	8.35	8.36	8.37	**8.53**
		4	8.02	8.45	8.52	8.55	8.60	**8.64**
Rear	X-axis	1	2.36	4.4	4.56	4.56	4.81	**4.85**
		2	9.45	9.81	9.81	9.82	9.82	**9.84**
		3	7.93	8.19	8.21	**8.22**	**8.22**	**8.22**
		4	8.04	8.34	8.47	8.50	8.54	**8.57**
	Y-axis	1	2.91	3.44	5.69	5.77	5.75	**6.37**
		2	10.7	11.0	11.0	11.0	11.0	**11.4**
		3	8.99	9.02	9.05	9.05	9.06	**9.90**
		4	8.45	8.69	8.76	8.81	8.82	**9.22**
	Z-axis	1	2.71	4.37	5.36	5.40	5.42	**5.74**
		2	10.3	10.5	10.6	**10.7**	**10.7**	**10.7**
		3	8.5	8.62	8.62	8.63	8.63	**9.01**
		4	7.48	7.73	7.79	7.82	7.84	**7.86**

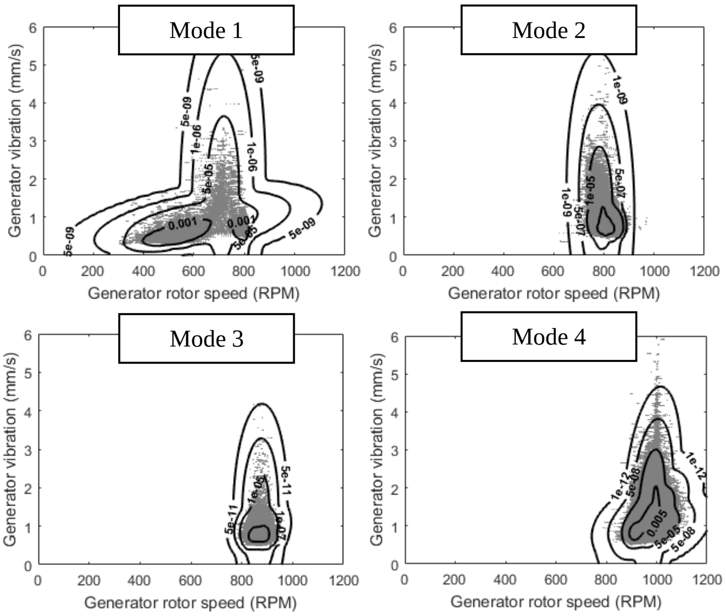


Figure 3.23 - Gaussian mixture modelling (where $K = 5$) applied to control mode clusters of generator vibration (X-axis, input side) against generator rotor speed

(Contour lines indicate outwardly decreasing levels of probability)

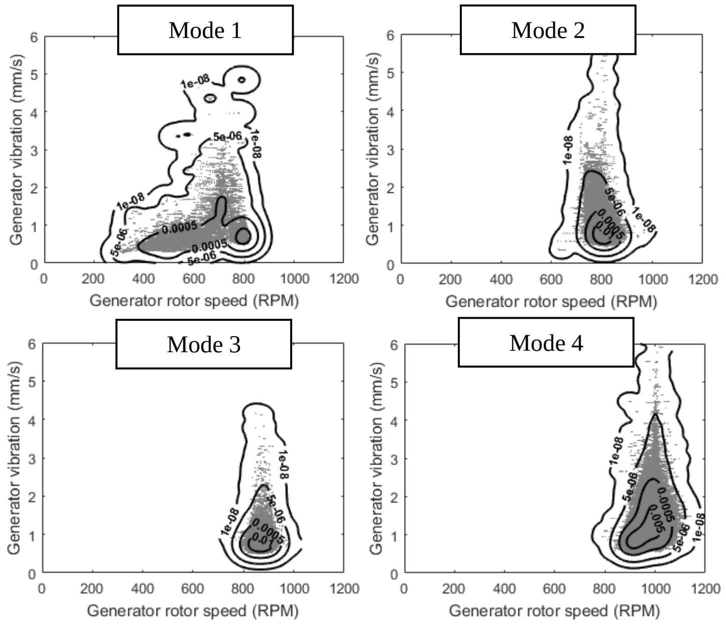


Figure 3.24 - Kernel density estimation applied to control mode clusters of generator vibration (X-axis, input side) against generator rotor speed

(Contour lines indicate outwardly decreasing levels of probability)

3.6.2.2 Anomaly Thresholds

Anomaly thresholds were set using each KDE model, where an outer probability value of T was selected to define a region of normality around each cluster of data, as in (3.16), p. 45. Data points with modelled probability values above this threshold would be defined as normal responses, and conversely, data points with modelled probability values below this threshold would be defined as anomalous.

Threshold values T were set as the 99% confidence interval (CI) of the KDE model's estimated distribution. These values are detailed in Table 3.8 for vibration parameters in each control mode.

Table 3.8 - 99% CI threshold values T for each generator vibration parameter KDE model

Generator side	Vibration Axis	Control mode	T
Input	X-axis	1	1.14×10^{-4}
		2	5.52×10^{-4}
		3	2.28×10^{-4}
		4	1.60×10^{-4}
	Y-axis	1	6.16×10^{-5}
		2	3.98×10^{-4}
		3	6.03×10^{-4}
		4	1.68×10^{-4}
	Z-axis	1	1.05×10^{-4}
		2	1.76×10^{-4}
		3	2.08×10^{-4}
		4	1.33×10^{-4}
Rear	X-axis	1	7.37×10^{-5}
		2	1.81×10^{-4}
		3	2.36×10^{-4}
		4	1.38×10^{-4}
	Y-axis	1	1.01×10^{-4}
		2	5.77×10^{-4}
		3	1.86×10^{-4}
		4	2.39×10^{-4}
	Z-axis	1	8.16×10^{-5}
		2	3.97×10^{-4}
		3	6.31×10^{-4}
		4	1.70×10^{-4}

3.6.3 Bearing Vibration

Data analysis revealed relationships between high speed bearing vibrations and the operational parameters generator rotor speed and output power. Closer examination of these relationships revealed that this data did not follow a relationship where $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{\theta})$, but varied continuously between zero and maximum values dependent on the rotation speed or output power of the turbine. This demonstrated a trend across each relationship's envelope, where peaks in the data's variance occurred at different levels of generator rotor speed and particularly output power.

Bearing vibration parameters were defined through a process referred to as envelope fitting (section 3.5.5.2, p. 55), where a curve was fitted to the maximum values of vibration sampled across the range of output power to give a model $f(x, \boldsymbol{\theta})$ where f is a parametric function with parameters $\boldsymbol{\theta}$ x is output power.

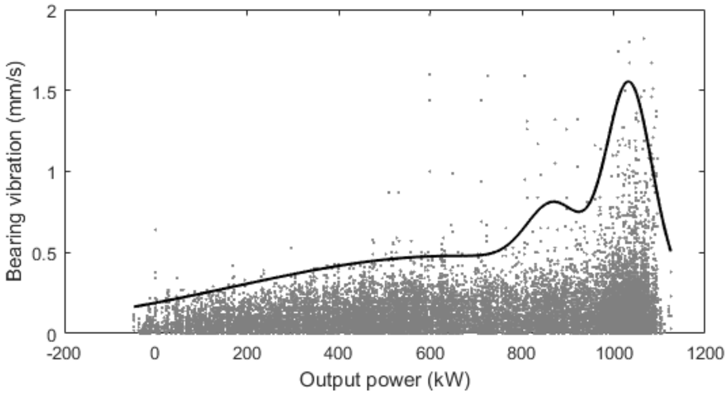
For these models, evaluation involved testing the accuracy of different parametric functions for each vibration parameter. Anomaly thresholds were then applied based on distances from the envelope model.

3.6.3.1 Model Selection

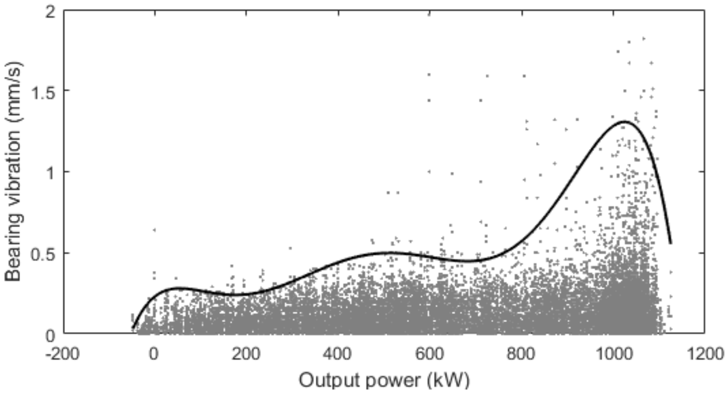
Models were trained using 70% of the available data, randomly sampled, and tested using the remaining 30%. A number of different functions were tested to find the best fit to the envelope of each vibration parameter against output power. These included Gaussian and polynomial functions of varying orders ((3.33) and (3.34), p. 58), to find which is most suitable for defining a series of peaks observed across each envelope, Figure 3.25.

RMSE and R^2 metrics ((3.35) and (3.36), p. 59) were used to assess how well each trained function modelled the test data, where functions with the lowest RMSE value and R^2 value closest to 1 are best fits.

Table 3.9 and Table 3.10 detail these results for fix and loose bearing vibration parameters respectively. Higher order Gaussian and polynomial functions were found to best fit fix bearing vibration parameters, reflective of the number of peaks across the relationship. Conversely, lower order Gaussian functions were found to best fit loose bearing vibration parameters, where a lower number of peaks formed across each relationship.



(a) 6th order Gaussian function



(b) 6th order polynomial function

Figure 3.25 – Envelope fitting on fix bearing vibration (X-axis) using different functions

Table 3.9 - Goodness of fit metrics for fix bearing vibration model test errors

Function	**Order**	**Fix Bearing (X-axis)**		**Fix Bearing (Y-axis)**		**Fix Bearing (Z-axis)**	
		RMSE	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
Gaussian	3rd	0.2356	0.7965	0.2786	0.6797	**0.1520**	0.8057
	4th	0.2238	0.8212	0.2530	0.7428	0.1728	0.7553
	5th	0.2287	0.8181	0.2408	0.7443	0.1525	0.8156
	6th	**0.2231**	**0.8316**	0.2949	0.6689	**0.1520**	**0.8195**
Polynomial	3rd	0.2783	0.7038	0.3139	0.5759	0.2403	0.4931
	4th	0.2466	0.7694	0.2639	0.7026	0.2405	0.4964
	5th	0.2447	0.7747	0.2564	0.7218	0.2331	0.5310
	6th	0.2280	0.8061	**0.2351**	**0.7679**	0.1819	0.7169

Table 3.10 - Goodness of fit metrics for loose bearing vibration model test errors

Function	**Order**	**Loose Bearing (X-axis)**		**Loose Bearing (Y-axis)**	
		RMSE	**R^2**	**RMSE**	**R^2**
Gaussian	3rd	0.1835	0.8741	**0.1816**	**0.7671**
	4th	**0.1759**	**0.8873**	0.2056	0.7091
	5th	0.1925	0.8684	0.1937	0.7486
	6th	0.1765	0.8739	0.1990	0.7421
Polynomial	3rd	0.2777	0.6990	0.2950	0.3589
	4th	0.2758	0.7056	0.2883	0.3929
	5th	0.2601	0.7403	0.2751	0.4517
	6th	0.2120	0.8290	0.2207	0.6500

3.6.3.2 Anomaly Thresholds

Thresholds were selected using a similar process to section 3.6.1.2, p, 60, based on the work in [85], where anomaly thresholds were applied based on the PDF of prediction errors $\boldsymbol{r}$. In this case however, the prediction error was the difference between each model and sampled envelope values 69. In addition, only an upper threshold $U(k)$ was applied, identifying anomalies lying above a certain distance from the bound. This allowed for some deviation from each model, reducing false alarm rates due to model fitting errors, noise and single point anomalies not reflective of an actual change in response within the system.

$$\boldsymbol{r} = \boldsymbol{y}_{\text{env}} - f(\boldsymbol{x}, \boldsymbol{\theta}) \tag{3.40}$$

Figure 3.26 illustrates this process, where the PDF of prediction errors $\boldsymbol{r}$ is estimated through kernel density estimation [59] and $U(k)$ is calculated as in (3.38), p. 60.

Table 3.11 lists the calculated upper threshold values $U(k)$ for each vibration parameter.

Table 3.11 - Upper threshold values $U(k)$ for each bearing vibration model

Component	**Vibration Axis**	$\boldsymbol{U(k)}$ **(mm/s)**
Fix bearing	X-axis	0.59
	Y-axis	0.68
	Z-axis	0.42
Loose bearing	X-axis	0.62
	Y-axis	0.50

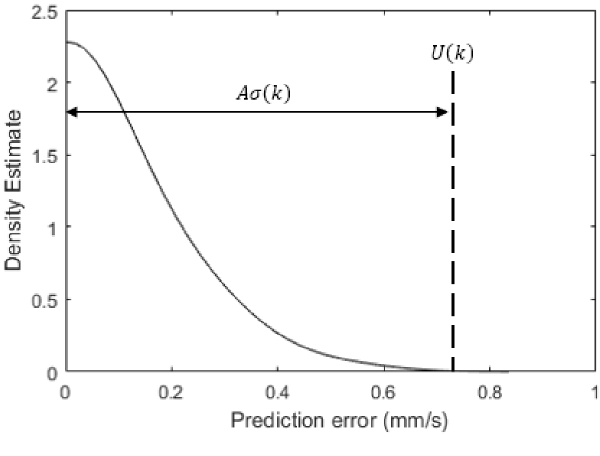


Figure 3.26 - Thresholds applied to kernel density estimation of model prediction error

3.7 Deployment

The final stage of the CRISP-DM methodology is deployment, where models are used in operation. At this stage, the knowledge gained and models constructed from data mining must be presented in a way that is useful to the user [46].

In this case, models are used to detect changes in the response of the turbine from normal behaviour, indicating a fault within the turbine to be examined by the operator.

This section details how each anomaly detection model constructed in this work can be used in operation. Each model was tested on the available operational data from the HS1000 to detect anomalies within this dataset. Anomalies detected by each model were examined in detail, indicating how an operator could determine the cause of anomalous behaviour from the data.

3.7.1 Gearbox Vibration

Gearbox vibration models were tested on their ability to identify anomalies under both healthy and simulated fault conditions.

Under normal conditions, models were able to detect an anomaly representative of a change in the turbine's control scheme. The models were also capable of identifying simulated faults representing a gear tooth crack and gear misalignment. Analysis of the prediction error in each case indicated how operators can use this modelling technique to interpret and distinguish between these different anomalous behaviours.

3.7.1.1 Fault Simulation

Data utilised for this study were recorded over a period of normal turbine operation, where no known examples of mechanical faults were present within the dataset. Therefore, to test the response of each model under fault conditions, data were altered to simulate the onset of a fault within the gearbox.

The effect of gearbox faults on vibration signatures are well known, with different failure modes producing specific changes in measured dynamics of vibration (a detailed analysis of gearbox failure modes and their effect on vibration is provided by [87]). In this study, two categories of fault were simulated: a cracked or broken gear tooth and gear misalignment. Fault signals $y_{\text{fault}}(t)$ were generated by adding simulated fault signatures $F(t)$ to measured vibration signals $y(t)$ as in (3.41), where $\epsilon(t)$ is Gaussian white noise.

$$y_{\text{fault}}(t) = y(t) + F(t) + \epsilon(t) \tag{3.41}$$

A cracked or broken gear tooth produces vibration impacts as the gear tooth comes into contact with the pinion. This produces a non-stationary vibration signature, as a series of periodic spikes or impulses occurring at the gear rotation frequency, f_{gear}. Mohammed, et al., use simulated vibration signals to show that as a gear tooth crack grows, the amplitude of vibration spikes increase [88]. This fault signal was simulated as in Figure 3.27, where vibration spikes are added to the vibration signal at periods of $\Delta = \frac{1}{f_{\text{gear}}}$.

Gear misalignment produces a stationary fault signature through an increase in vibration at harmonics of the gear mesh frequency, f_{mesh}, containing sidebands of the gear rotation frequency, f_{gear}, due to amplitude modulation [87]. This was simulated as in (3.42), giving N harmonics of gear mesh frequency with sidebands at $\pm f_{\text{gear}}$, where A_i and B_i are the amplitudes of gear mesh and sideband components respectively. Figure 3.28 shows the fault signature for the first 3 harmonics of gear mesh frequency.

$$\begin{aligned} F(t) = \sum_{i=1}^{N} A_i \sin 2\pi i f_{\text{mesh}} t + B_i \sin 2\pi i \left(f_{\text{mesh}} - f_{\text{gear}}\right) t \\ + B_i \sin 2\pi i \left(f_{\text{mesh}} + f_{\text{gear}}\right) t \end{aligned} \tag{3.42}$$

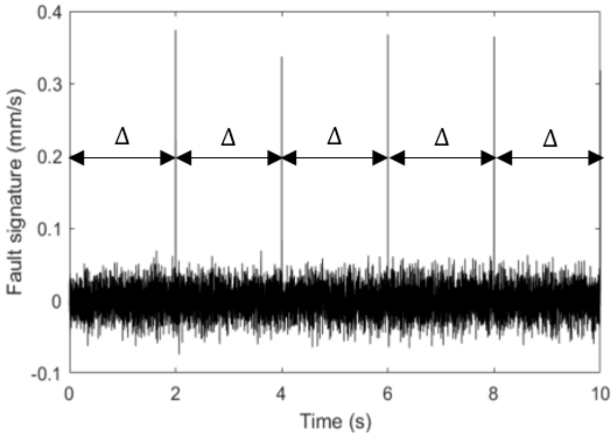


Figure 3.27 - Simulated gear tooth crack fault vibration signature

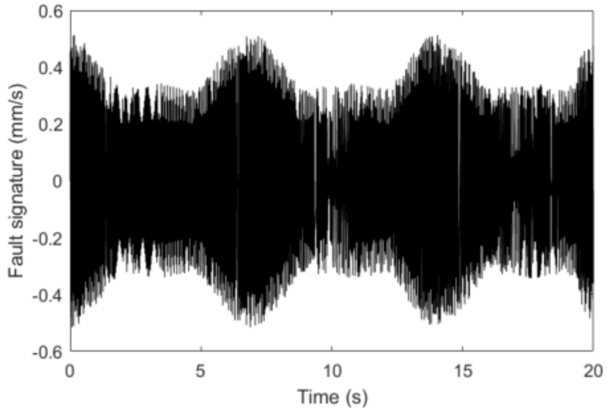


Figure 3.28 - Simulated gear misalignment fault vibration signature

3.7.1.2 Results

Each gearbox vibration model was evaluated under both normal and faulty conditions. Normal conditions were reflected through data recorded from the HS1000 turbine operating over a period from between September 2013 to February 2014. Faulty conditions were reflected through simulated fault data, as described above.

Firstly, Figure 3.29 shows the averaged prediction errors for X-axis gearbox vibration from the available operational data. Prediction errors were observed to cross the upper anomaly threshold during a single tidal cycle. This reveals a period of operation during which the pitch and torque control scheme was altered briefly and does not represent a fault within the machine. The previous control scheme was restored shortly after the introduction of the alternative, and the prediction error returned to its previous range of values.

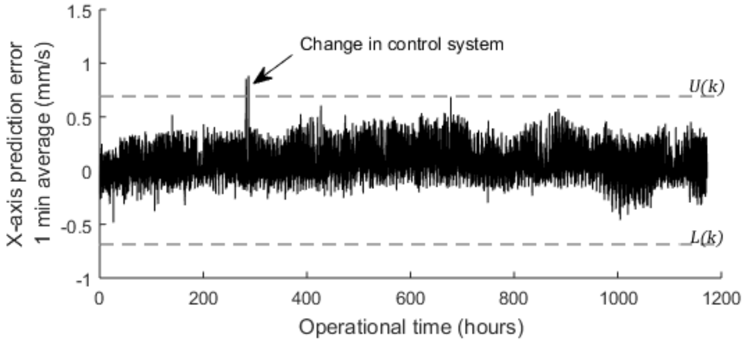
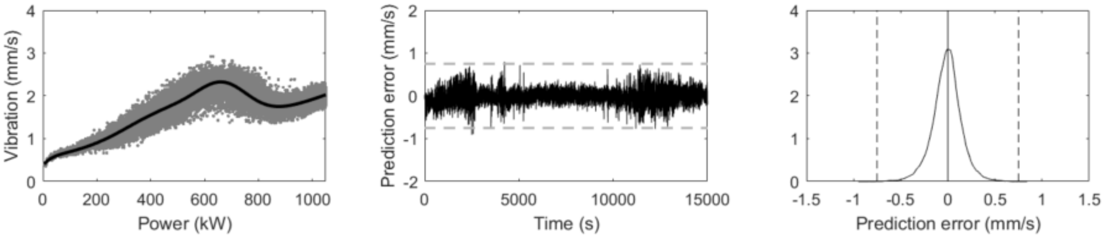


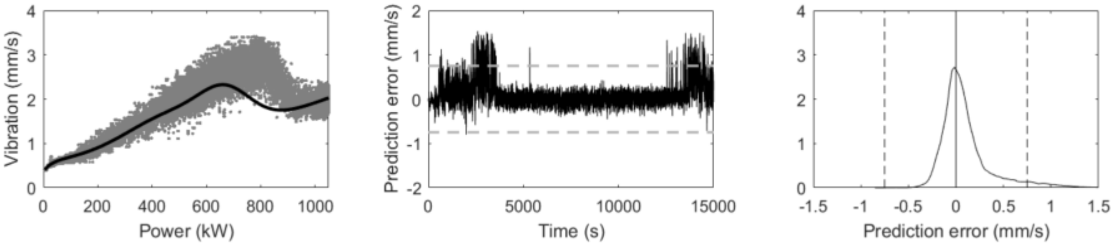
Figure 3.29 - Prediction error of healthy gearbox behaviour, with a change in control system identified

The response of the gearbox vibration model from each machine condition over a single tidal cycle, including normal behaviour, alternative pitch/torque control scheme, simulated gear tooth crack and simulated gear misalignment, are shown in Figure 3.30 to Figure 3.33 respectively. Each figure shows the gearbox vibration model, the prediction error over time and the PDF of the prediction error. These plots can be inspected by operators to understand why the anomalous behaviour has been detected.



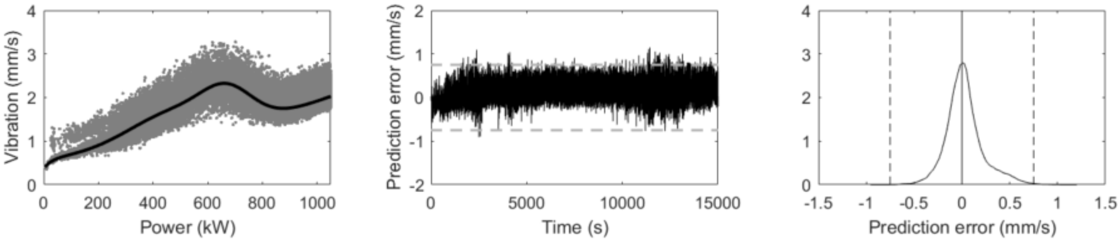
(a) (b) (c)

Figure 3.30 - Normal response (a) measured vibration against model, (b) prediction error over time and (c) prediction error PDF



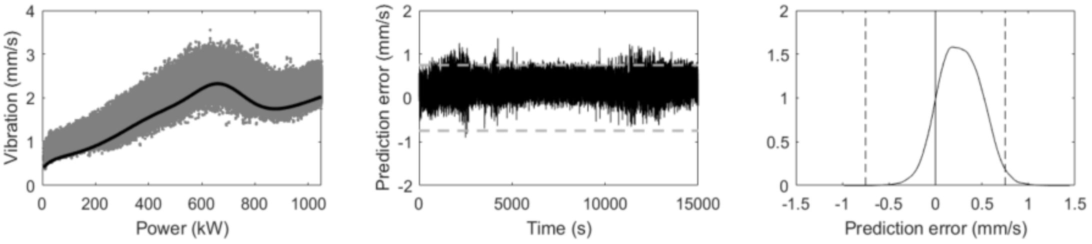
(a) (b) (c)

Figure 3.31 - Control scheme change (a) measured vibration against model, (b) prediction error over time and (c) prediction error PDF



(a) (b) (c)

Figure 3.32 - Simulated gear tooth crack (a) measured vibration against model, (b) prediction error over time and (c) prediction error PDF



(a) (b) (c)

Figure 3.33 - Simulated gear misalignment (a) measured vibration against model, (b) prediction error over time and (c) prediction error PDF

Figure 3.30 shows the vibration model and prediction errors during normal machine conditions. Vibration values follow the trained model closely, with prediction errors remaining within the thresholds of normal behaviour. The PDF of the prediction error is symmetrical around zero.

Figure 3.31 shows the vibration model and prediction errors during the pitch/torque control scheme change identified in Figure 3.29. The effect of the control scheme on the response of gearbox vibration is discussed in section 3.3.2.6, p. 38, where changes in control mode cause maxima and minima across the relationship of gearbox vibration and output power. Here, a change in the control scheme has caused peak vibration to change from around 700 kW to 800 kW, and the subsequent minima to shift from around 900 kW closer to 1000 kW. This therefore causes levels of vibration to increase above expected levels as the turbine approaches rated power, causing anomalies during this period of the tidal cycle. However, the shape of the relationship remains the same, Figure 3.31(a), and the prediction error decreases to normal levels mid tidal cycle, Figure 3.31(b), as rotation speed and output power increase. This would indicate to an operator that the anomaly is related to an operational change and not due to a mechanical gearbox fault.

Figure 3.32 shows the vibration and prediction errors during a simulated gear tooth fault, where periodic impacts were added to the vibration signal at the gear rotation frequency as in (3.40). In this case, maxima and minima of the relationship remain, however levels of vibration increase across the relationship, Figure 3.32(a). This causes prediction errors to increase above the upper threshold consistently in time over the tidal cycle, Figure 3.32(b). The maximum value in the PDF remains at zero, however the PDF becomes positively skewed as levels of vibration increase. These would indicate to the operator that anomalies are related to component wear or a mechanical fault within the gearbox.

Finally, Figure 3.33 shows the vibration and prediction errors during simulated gear imbalance, simulated as in (3.42) by adding harmonics of gear mesh frequency components with sidebands of the gear rotation frequency to vibration signals. This fault causes increases in vibration and prediction errors similar to the gear tooth crack example, as the fault causes vibration to increase over a range of output power values, Figure 3.33(a) and (b). The PDF plot, Figure 3.33(c), also differs between faults, showing increased variance and a peak value shifted above zero for misalignment. These factors also indicate to an operator that anomalies are related to a mechanical fault within the gearbox and not an operational change.

3.7.2 Generator Vibration

The deployment of models for each generator vibration parameter and each control mode was tested using data from normal operating conditions, measured during a period of HS1000 turbine operation from September 2013 to February 2014.

In this case, generator vibration models detected an anomaly caused by a non-mechanical fault from the pitch controller.

3.7.2.1 Results

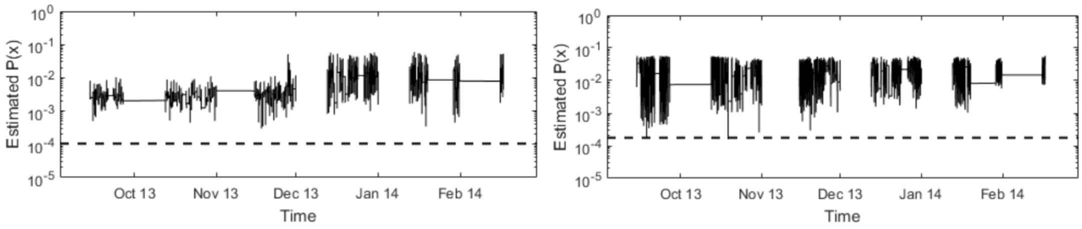
Figure 3.34 shows the predicted probability $\hat{P}(\boldsymbol{x})$ for each data point in each control mode cluster for the X-axis of the input side generator vibration. An anomaly was detected as probability values crossing below the anomaly threshold during one tidal cycle, in control modes 3 and 4. Similar results were observed across all generator vibration parameters.

This anomaly occurred during a tidal cycle where the turbine experienced a non-mechanical fault with the blade pitch controller system. Figure 3.35 and Figure 3.36 explore this anomaly in more detail, showing the generator rotor speed and generator vibration (X-axis input side) over time respectively.

In Figure 3.35, the turbine changes from a non-braking idle state (of rotation speeds around 100 RPM) to an operational state at approximately 14:50, accelerating to rotation speeds around 800 RPM. However, after approximately 7 minutes in operation, the turbine does not accelerate to rated values and instead rapidly decelerates to rotation speeds of around 300 RPM. Consultation with engineers at Andritz Hydro Hammerfest revealed this behaviour occurred due to a fault in the pitch controller.

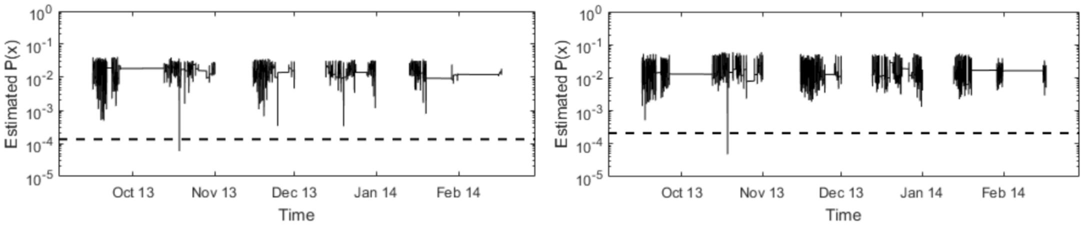
The moment of the vibration anomaly can be seen in Figure 3.36 where a large spike in vibration can be observed before the deceleration, as the turbine begins to transition to control modes 3 and 4. Vibration levels around this spike indicate the moment the fault becomes active, as pitch control is activated during these control modes, to limit the rotation speed of the machine to rated values as tidal flow increases (pitch control is discussed in detail in chapter 2). Following the activation of the fault, the turbine rapidly decelerates to a new state and vibration levels decrease to levels reflective of the generator rotating at 300 RPM.

This example shows the effectiveness of this modelling technique, as vibration models were able to correctly detect a faulty change in behaviour. In addition, by modelling data from each control mode separately, the source of this anomaly was able to be examined in more detail, as modes of operation are indicative of different control methods, e.g. torque or pitch control.



(a) Mode 1

(b) Mode 2



(a) Mode 3

(b) Mode 4

Figure 3.34 - Estimated $p(x)$ over time for generator vibration (X-axis, input side) in each control mode, where the dashed line indicates the probability threshold T

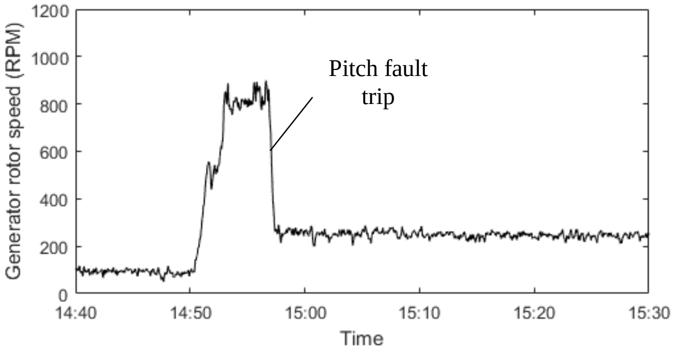


Figure 3.35 - Generator rotor speed over time during anomaly

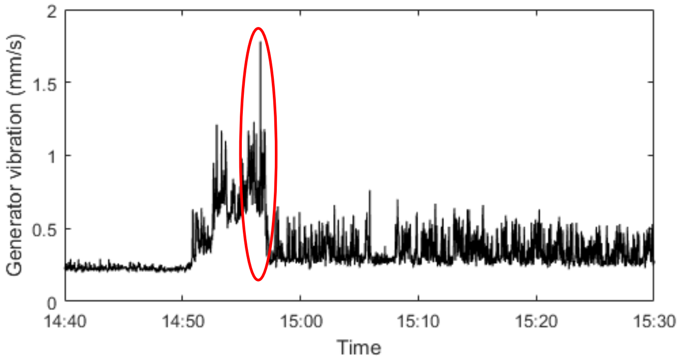


Figure 3.36 - Generator vibration (X-axis, input side) over time during anomaly

3.7.3 Bearing Vibration

The deployment of models for each bearing vibration parameter was also tested using data from normal operating conditions, measured during a period of HS1000 turbine operation from September 2013 to February 2014.

In this case, models detected an anomaly during shutdown caused by turbulent weather conditions, but not representative of faulty behaviour.

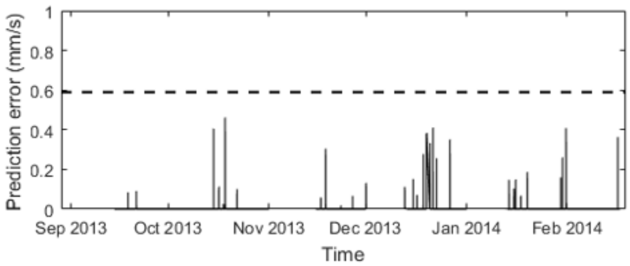
3.7.3.1 Results

Figure 3.37(a) to (e) show the response of each model over the period of the dataset, detailing the prediction error of bearing vibration data points from each corresponding envelope model. In each case, anomalies are detected above thresholds, as calculated in section 3.6.3.2, p. 69, indicated by a dashed line on each plot.

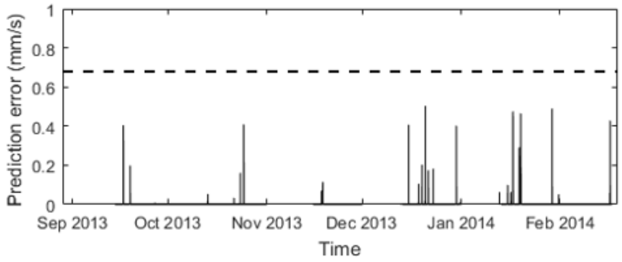
One anomalous response was detected within the dataset, from the loose bearing Y-axis vibration model during a single tidal cycle. Consultation with engineers at Andritz Hydro Hammerfest indicated this anomaly was triggered during a period of adverse weather, where high ocean waves triggered an unscheduled shutdown mid cycle.

Figure 3.38, Figure 3.39 and Figure 3.40 examine this anomalous response in greater detail. Figure 3.38 shows the measured data and model during the tidal cycle where the anomaly was detected, where a small number of anomalous data points can be observed at low output power of around 30 kW. This indicates the anomalies occurred during a turbine start-up or shutdown.

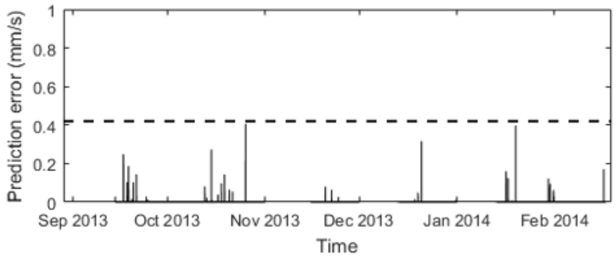
Figure 3.39 and Figure 3.40 show the prediction error of data points from the model and generator rotor speed over time during the anomaly respectively. A spike in vibration can be observed in Figure 3.39, triggering the anomaly. This occurs during the shutdown itself, as the turbine rapidly decelerates from around 1000 RPM to an idle state. This high spike in vibration behaviour can be expected due to high loading from an emergency shutdown, particularly during turbulent conditions, and is not linked to a fault within the turbine.



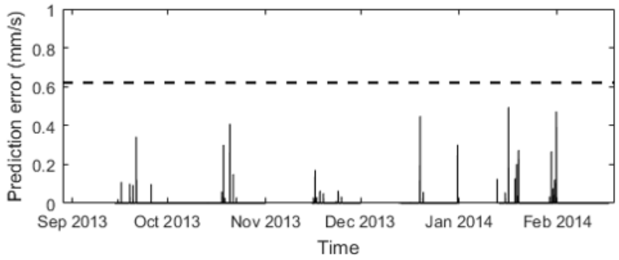
(a) Fix bearing X-axis



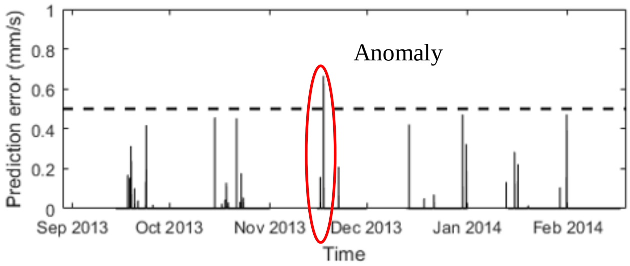
(b) Fix bearing Z-axis



(c) Fix bearing Z-axis



(d) Loose bearing X-axis



(e) Loose bearing Y-axis

Figure 3.37 – Prediction error between data and envelope mode, where dashed line indicates anomaly threshold

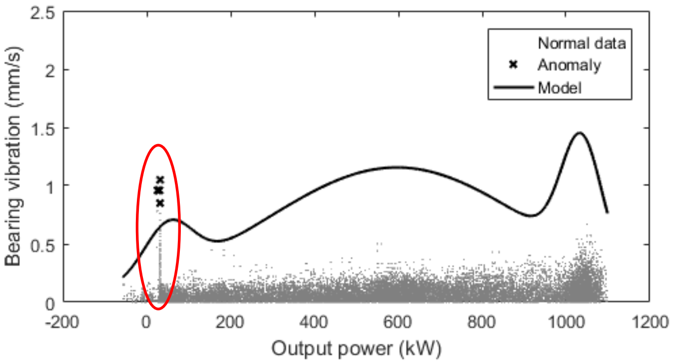


Figure 3.38 - Bearing vibration model (loose bearing Y-axis) during anomaly

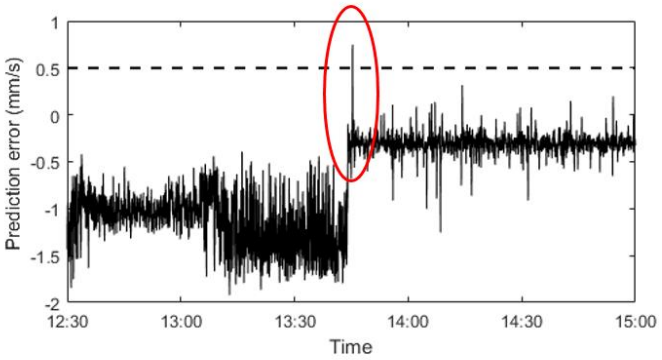


Figure 3.39 - Vibration data prediction error from model boundary (loose bearing Y-axis) over time during anomaly

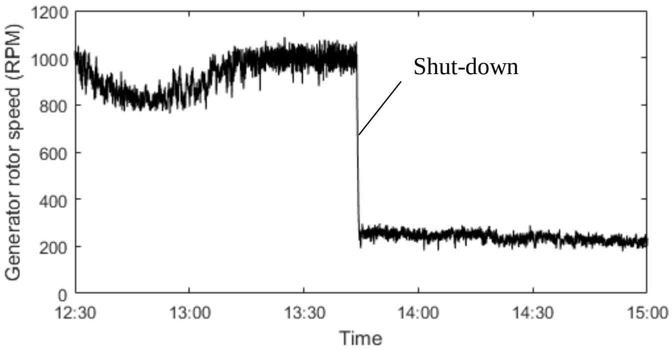


Figure 3.40 -Generator rotor speed over time during anomaly

3.8 Discussion

In this chapter, the CRISP-DM methodology was utilised to explore HS1000 tidal turbine condition data and construct models of normal behaviour for anomaly detection.

Within this process, data analysis is used to reveal key trends and relationships within a dataset, indicative of the normal response of the turbine over different modes of operation. These results are then modelled using anomaly detection techniques, where faults can be identified as the response of the system changes away from the modelled normal behaviour.

In this work, models of normal behaviour were defined for the HS1000 turbine's gearbox, generator and high speed bearings. Here it was found that different types of modelling technique were found to be most effective for different forms of vibration relationship observed across the turbine. In general, reconstruction based techniques (such as curve fitting) are most suitable for modelling gearbox vibrations, where vibration varies directly in response to rotation speed and output power. Here, the choice of function should be dependent on how the amplitudes of vibration vary in the response to the turbine's control scheme. Probabilistic techniques are best suited to modelling generator vibration, where data tends to cluster into areas dependent on the control mode of the turbine. Finally, envelope fitting techniques are most suitable in modelling bearing vibration, where data varies continuously between zero and a maximum amplitude dependent on the output power or rotation speed of the machine.

Such models were then trained and applied to past operational data from the turbine. These models were able to detect anomalous events from a change in control scheme, a non-mechanical pitch controller fault and a mid-cycle shutdown due to bad weather. These events however were not linked to mechanical damage of the turbine's components requiring additional inspections or maintenance.

Although this framework is successful in detecting anomalous behaviour, data from each event requires further analysis by turbine operators to determine the cause of an anomaly. This is reflected Figure 3.41, where the CRISP-DM methodology has been adapted for newly commissioned machinery with no past fault cases. Here, the process can return to the data understanding stage if an anomaly is detected during deployment, in order to investigate the root cause of the anomaly.

This process is best suited in new applications where there are few active assets with limited operational history, allowing operators to learn from the normal behaviour of machinery and automatically detect faults as they occur in operation. The remaining chapters in this thesis however examine how this methodology can be expanded for diagnostic and prognostic modelling as the time in operational of devices increase, allowing an intelligent condition monitoring system to learn from faults and failures of equipment after they are detected.

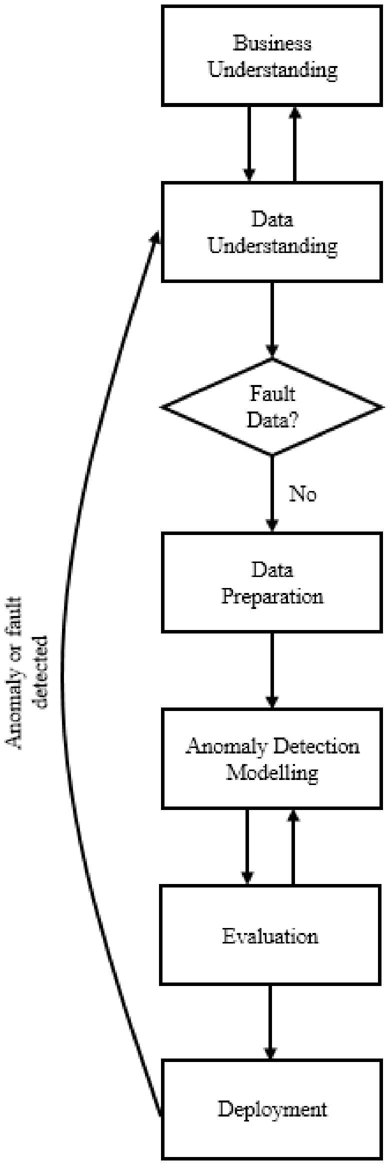


Figure 3.41 - CRISP-DM methodology for anomaly detection during initial deployment period

Chapter 4

4 Diagnostic Modelling

Following anomaly detection, diagnostics is the location and classification of a fault once it has been detected within the system. This provides more detailed information, allowing operators to more effectively plan maintenance.

Diagnostics can be performed by experienced engineers with thorough knowledge of the system and its response to failure through manual inspection of sensor data. However, the aim of an intelligent condition monitoring system is to perform diagnoses at an early stage from sensor data automatically, before faults develop and propagate through machinery.

In order to train diagnostic models, fault data from the system is required. Many current diagnostic techniques use feature-based methods [29], where multiple signal processing techniques are used to isolate and extract specific fault signals from data before a classification stage. However, deep learning, a recently advancing field of machine learning, offers an alternative to feature-based methods, where features are extracted from data automatically using multi-layered neural network architectures [89].

This chapter details the fault diagnostic process for rotating machines. First, common failure modes of drive train components (gearboxes, bearings and generators) are described. Secondly, different feature-based diagnostic methods are detailed. Deep learning for rotating machine diagnostics is then described. Two case studies are presented, describing a feature-based method and deep learning method for diagnosing a generator fault on the HS1000 and multiple faults from a gearbox test rig. Finally, the integration of diagnostic models into intelligent condition monitoring systems for new applications is discussed.

4.1 Rotating Machine Fault Modes and Vibration Characteristics

Vibration is one of the most effective diagnostic indicators for faults within rotating equipment [90]. One reason vibration is effective is because of the immediacy with which warnings of failure are provided. Vibration typically responds earlier than other fault indicators such as oil quality, acoustic emissions or temperature, up to several months before complete failure, as shown in Figure 4.1 [20].

In addition to giving an early indication of faults, vibration data also provides enough information for different rotating machine faults to be distinguished from one another [90], making it a suitable measurement for constructing diagnostic models.

Vibration can be measured in three separate quantities: displacement, velocity and acceleration. The choice of vibration measurement usually relates to frequency range, with displacement most suitable for low frequency vibration (up to 10s of Hz range), velocity most suitable for mid frequency vibration (up to 100s of Hz) and acceleration for high frequency vibration (up to kHz range) [91]. High resolution accelerometer sensors are therefore most suitable for rotating machine diagnostics, giving the maximum range of frequencies to infer diagnostic information from (e.g. from low frequency rotations to high frequency natural frequencies).

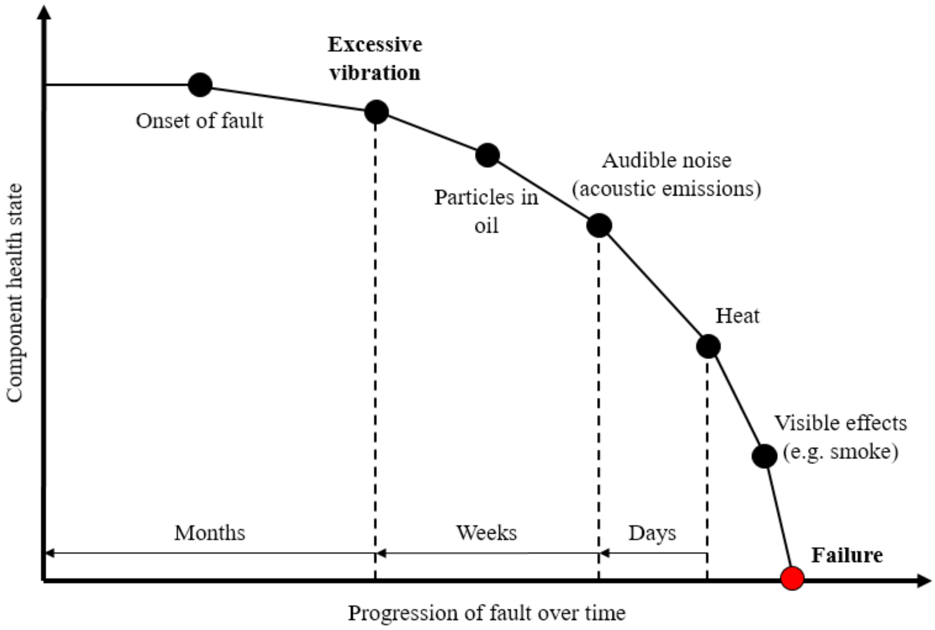


Figure 4.1 - Response of fault indicators as fault progresses over time, adapted from [20]

This section describes how vibration data can be used to diagnose faults within gearboxes, bearings and generators, detailing the characteristic vibration responses specific to fault modes from each component.

4.1.1 Gearbox Fault Modes

Gearboxes are used within turbines to translate slow rotational speeds of the rotor to fast rotational speeds within the generator and dissipate high loading from the driving input force (e.g. tidal flow) [45]. This is typically achieved through a number of stages to reduce the loading applied to individual gears. The HS1000 tidal turbine contains three gear stages: one planetary gear stage followed by two parallel axis gear stages.

Gearbox faults can either occur due to damage to gears or misalignment between gear stages [92] [93]. This includes:

- Gear tooth crack/breakage – spalling, cracks or breakages within single gear teeth caused by excessive loading or loss of lubrication
- Gear tooth wear – gear teeth are worn more evenly across the surface of the teeth due to aging, excessive loading or lack of lubrication
- Gear misalignment – gear stages are not aligned in parallel causing radial movement

4.1.1.1 Characteristic Frequencies

Gearbox faults can be identified from vibration data in the frequency spectrum through a series of characteristic frequency values, commonly gear rotation, mesh and hunting tooth frequencies, as well as accompanying harmonics and sidebands [92].

The gear mesh frequency [94] for parallel axis gear stages is shown in (4.1), where n_{gear} and n_{pinion} are the number of teeth in the gear and pinion respectively and f_{gear} and f_{pinion} are the rotation frequencies of the gear and pinion respectively.

$$f_{\text{mesh}} = n_{\text{gear}} f_{\text{gear}} = n_{\text{pinion}} f_{\text{pinion}} \tag{4.1}$$

Characteristic frequencies within planetary gearbox stages [94] are more complex, shown in equations (4.2) to (4.5), where N_{P}, N_{R} and N_{S} are the number of teeth in the planet, ring and sun gears respectively, M_{P} is the number of planet gears and f_{s} is the rotation frequency of the sun gear.

$$f_{\text{planet}} = \frac{(N_{\text{P}} - N_{\text{R}})N_{\text{s}}}{(N_{\text{R}} + N_{\text{S}})N_{\text{P}}} f_{\text{S}} \tag{4.2}$$

$$f_{\text{carrier}} = \frac{N_{\text{S}}}{N_{\text{R}} + N_{\text{S}}} f_{\text{S}} \tag{4.3}$$

$$f_{\text{planet-pass}} = \frac{N_{\text{S}} M_{\text{P}}}{N_{\text{R}} + N_{\text{S}}} + f_{\text{S}} \tag{4.4}$$

$$f_{\text{planet-mesh}} = \frac{N_{\text{R}} N_{\text{S}}}{N_{\text{R}} + N_{\text{S}}} f_{\text{S}} \tag{4.5}$$

The hunting tooth frequency [92] occurs when two damaged teeth mesh together. This value is usually of a very low frequency for gear stages as meshing of two specific teeth happens only once in every few cycles. This frequency is calculated for each gear stage as in (4.6), where n_{gear} and n_{pinion} are the number of teeth in the gear and pinion respectively, f_{gear} is the rotation frequency of the gear and G is the highest common factor between the number of gear teeth and pinion teeth.

$$f_{\text{hunting tooth}} = \frac{f_{\text{gear}} G}{n_{\text{gear}} n_{\text{pinion}}} \tag{4.6}$$

4.1.1.2 Amplitude Modulation and Sidebands

Amplitude modulation is a clear indicator of faults within rotating equipment [92], particularly for gear damage. This occurs when the amplitude of a particular frequency component f_{c} oscillates at a lower frequency f_{a}, Figure 4.2. This effect introduces additional components within the frequency spectrum at either side of the frequency component occurring at ± the frequency of the amplitude oscillation, known as sidebands, Figure 4.3.

Figure 4.2 - Amplitude modulation frequency components

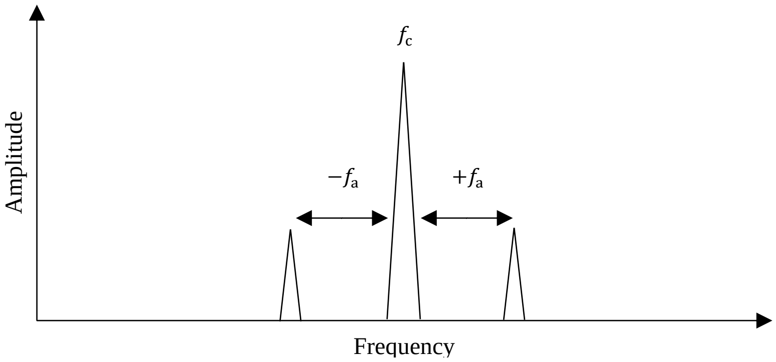


Figure 4.3 - Amplitude modulation in the frequency spectrum

Amplitude modulation commonly causes the amplitude of frequency components to oscillate with the rotation speed, causing sidebands of $\pm f_{\text{R}}$, the rotation frequency, around characteristic frequencies. The effect of amplitude modulation increases as damage/wear increases, increasing the amplitude of sidebands but not necessarily the amplitude at the characteristic frequency [92]. Monitoring the amplitude of sidebands as well as amplitudes of characteristic frequencies is therefore useful for identifying and diagnosing the severity of gear faults.

4.1.1.3 Normal Response

A healthy gearbox, containing no damage or significant wear, will contain characteristic frequencies at a low amplitude, as shown in Figure 4.4. Vibration components can be observed at the gear and pinion rotation frequencies (f_{gear} and f_{pinion}) which may also produce lower amplitude harmonics. Gear mesh frequencies (f_{mesh}) may also be observed, with sidebands at $\pm f_{\text{gear}}$.

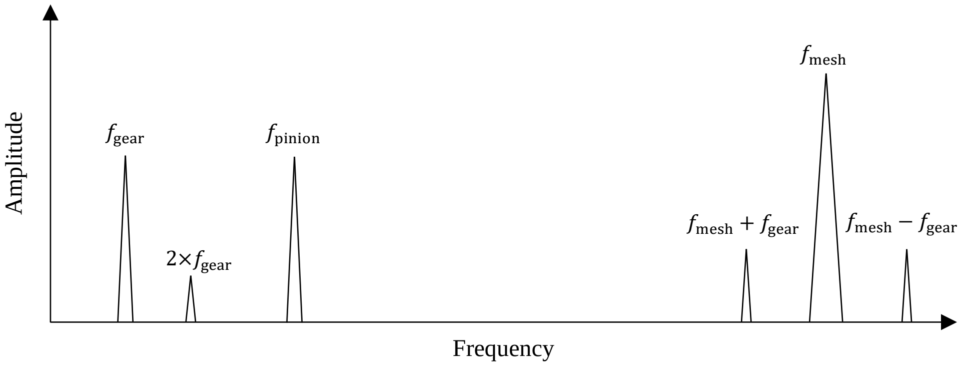


Figure 4.4 - Frequency response of a healthy gearbox, adapted from [92]

4.1.1.4 Gear Tooth Wear

Gear tooth wear will increase the effect of amplitude modulation of the gear mesh, increasing sidebands around f_{mesh} at $\pm f_{\text{gear}}$ [92]. However, the amplitude of f_{mesh} may or may not increase significantly as sidebands increase. The natural frequency of the gear (f_{natural}), related to the structural resonance of the gear, may also appear on the frequency spectrum, also containing sidebands at $\pm f_{\text{gear}}$. Here, a natural frequency is excited as the impact forces of individual gear teeth meshes increase. The exact value of the natural gear frequency may not be known before a fault occurs. These effects are shown in Figure 4.5.

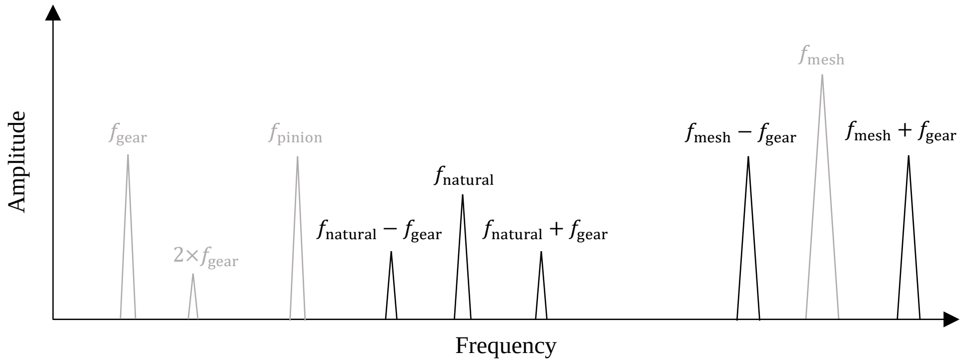


Figure 4.5 - Frequency response of a gearbox with tooth wear, adapted from [92]

4.1.1.5 Cracked or Broken Tooth

An individual cracked or broken tooth will cause an increase in amplitude at the gear rotation frequency, f_{gear} [92], Figure 4.6. This may also be observed in the raw time-based signal, where spikes in vibration occur at intervals of $\Delta = \frac{1}{f_{\text{gear}}}$, Figure 4.7.

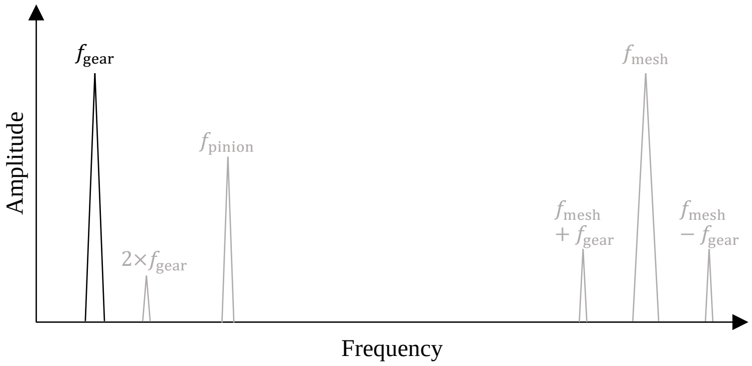


Figure 4.6 - Frequency response of broken or cracked tooth, adapted from [92]

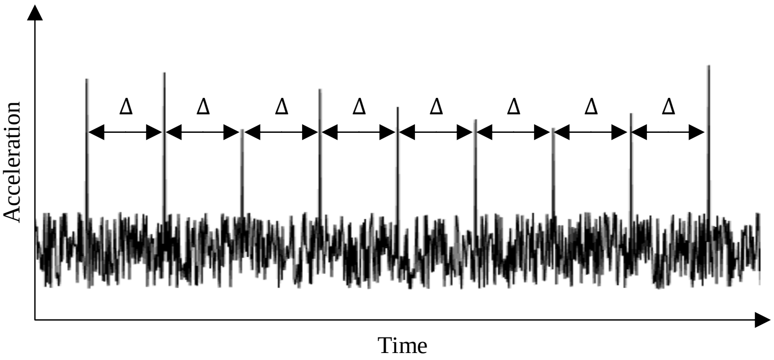


Figure 4.7 - Time-based signal of broken or cracked tooth with periodic impacts at Δ

4.1.1.6 Gear Misalignment

Misalignment within gear stages will cause harmonics of the gear mesh frequency to appear in the spectrum, with particular high amplitudes at 2x or 3x f_{mesh} [92]. Harmonics may also be accompanied by sidebands at $\pm f_{\text{gear}}$. This effect is shown in Figure 4.8.

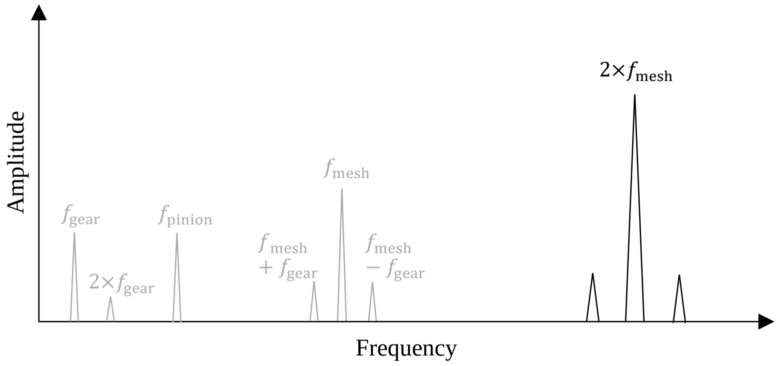


Figure 4.8 - Frequency response of gear misalignment, adapted from [92]

4.1.1.7 Hunting Tooth

Hunting tooth occurs when two damaged teeth (one from the gear and one from the pinion) come into contact during rotation [92]. This produces a low frequency vibration component, possibly accompanied by harmonics, Figure 4.9.

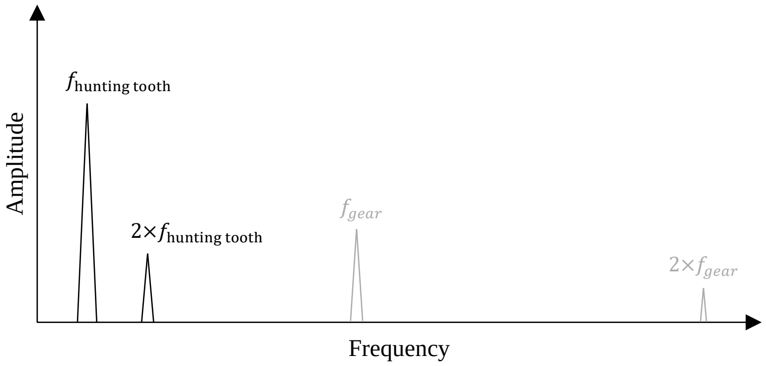


Figure 4.9 - Frequency response of hunting tooth, adapted from [92]

4.1.2 Bearing Fault Modes

Bearings are designed to support the rotation of shafts and constrict non-rotational movement [95]. Within tidal turbines, bearings are located to support shafts from within the rotor, gearbox and generator, constraining axial and radial displacement.

Figure 4.10 shows the typical components within a rolling element bearing: outer raceway, inner raceway cage and rolling elements. Rolling elements can either be spherical balls, cylindrical rollers or tapered cylindrical rollers [96]. When supporting a rotating shaft within a turbine, the outer raceway is typically fixed to the supporting structure (e.g. gearbox, generator, etc.) with the inner raceway rotating with the shaft.

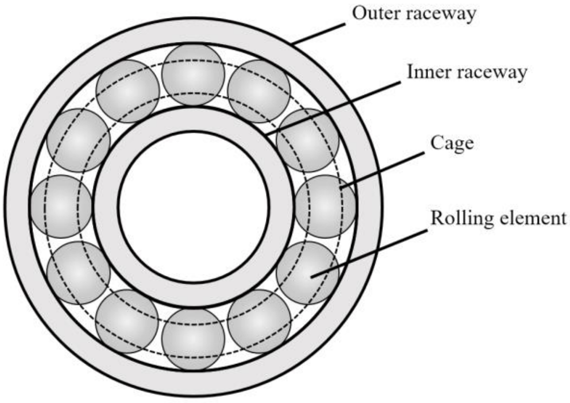


Figure 4.10 - Rolling element bearing components, adapted from [96]

Bearing faults can occur due to fatigue, wear, excessive loading, improper installation, lubrication problems or manufacturing defects [97] [98]. This can lead to damage to bearing components in the following ways:

- Spalling – Damage to the inner or outer raceways in the form of cracking or flaking [97] [98].
- Brinelling – Indentation in raceways caused by pressure from rolling elements [97] [98].
- Deformation – Raceways or cage bending or changing geometry due to loading or excessive temperature [97].
- Smearing – Metal being removed from one component and collecting on another (e.g. metal gathering at the ends of cylindrical rollers due from raceways to excessive loading or loss of lubrication) [97].
- Corrosion – Bearing materials become discoloured and damaged due to the factors from the surrounding environment [97] [98].
- Contamination – Dirt, dust or debris finding its way inside bearings causing damage (indentation, scratches, etc.) to raceways and rolling elements [98].
- Misalignment – Alignment of raceways and rolling elements becomes non-parallel [98].

4.1.2.1 Bearing Characteristic Fault Frequencies

Bearing faults are traditionally identified through a series of fundamental fault frequencies: fundamental train frequency (FTF), ball spin frequency (BSF), ball pass frequency of the outer ring (BPFO) and ball pass frequency of the inner ring (BPFI). These frequencies are dependent on the dimensions of the bearing and its rotation speed, calculated as in (4.8) to (4.11) [99], where f_{R} is the rotation frequency in Hz or rotations per second, Bd is the ball diameter in mm, Pd is the pitch diameter in mm, $\theta_{\mathrm{contact}}$ is the contact angle in degrees and Nb is the number of rolling elements (Figure 4.11).

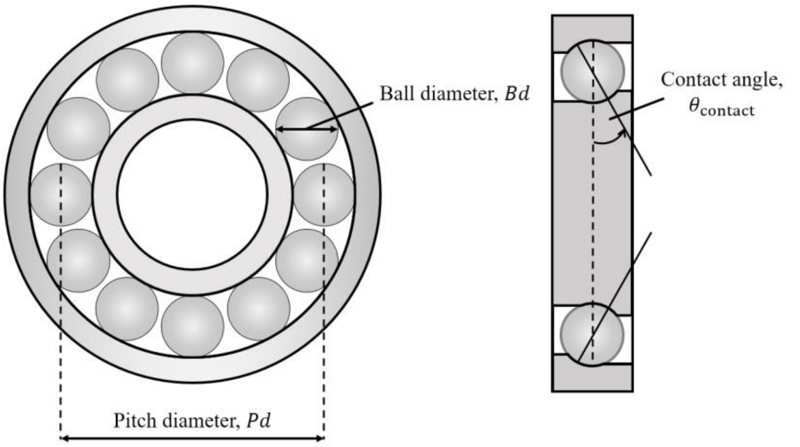


Figure 4.11 - Rolling element bearing specifications required to calculate fault frequencies

$$FTF = \frac{f_{\mathrm{R}}}{2}\left(1 - \frac{Bd}{Pd}\cos(\theta_{\mathrm{contact}})\right)\ (\mathrm{Hz}) \tag{4.8}$$

$$BSF = \frac{Pd}{2Bd} f_{\mathrm{R}}\left(1 - \left(\frac{Bd}{Pd}\right)^2 \cos(\theta_{\mathrm{contact}})^2\right)\ (\mathrm{Hz}) \tag{4.9}$$

$$BPFO = \frac{Nb}{2} f_{\mathrm{R}}\left(1 - \frac{Bd}{Pd}\cos(\theta_{\mathrm{contact}})\right)\ (\mathrm{Hz}) \tag{4.10}$$

$$BPFI = \frac{Nb}{2} f_{\mathrm{R}}\left(1 + \frac{Bd}{Pd}\cos(\theta_{\mathrm{contact}})\right)\ (\mathrm{Hz}) \tag{4.11}$$

As well as exciting vibrations at these fundemental frequencies, bearing failure can also produce vibrations at natural resonant frequencies. These frequency often are of a high value, between 3 and 50 kHz [99] and occur as rolling elements periodically pass over cracked or damaged areas of the bearing. Conventional approaches, such as the FFT, may not fully capture the frequency of these periodic impacts. The envelope of the vibration signature is therefore commonly used [92], providing a better representation of the frequency of bearing impacts. This is represented in Figure 4.12, Figure 4.13 and Figure 4.14, where the difference between the frequency spectrum of the raw vibration signal and envelope signal are compared.

Defects in the outer and inner raceways are the most common type of bearing fault [99]. This will produce frequencies at $BPFO$ and $BPFI$ respectively, as well as multiple harmonics within the envelope signal. Sidebands of f_{R} may also be present if the fault is severe. Defects within the rolling elements are less common, producing high amplitudes at the ball spin frequency BSF.

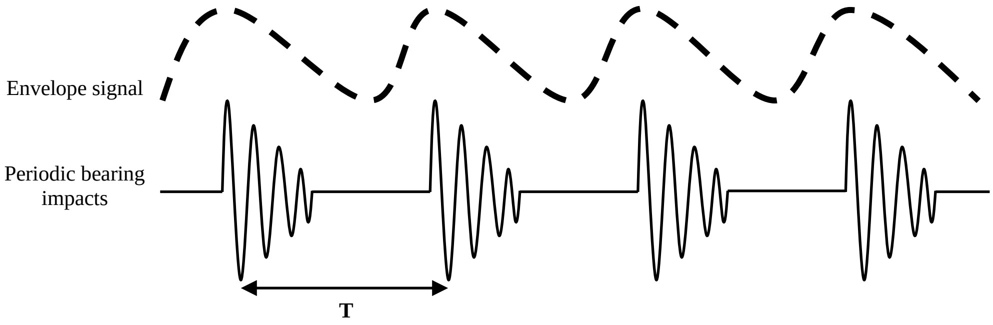


Figure 4.12 - Bearing fault envelope detection

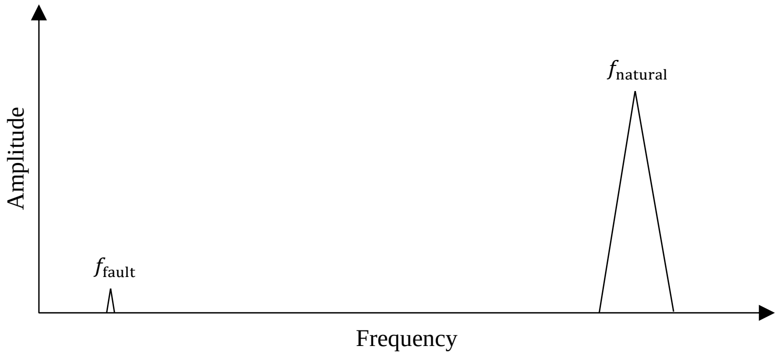


Figure 4.13 - Frequency spectrum of periodic impacts before envelope detection

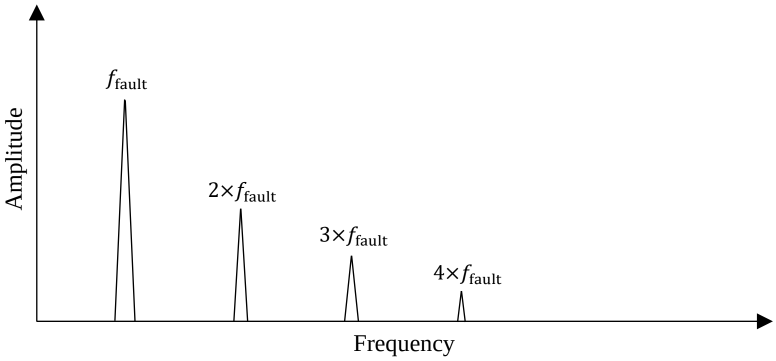


Figure 4.14 - Frequency spectrum of periodic impacts after envelope detection

4.1.3 Generator Fault Modes

The HS1000 tidal turbine contains a doubly-fed induction generator (DFIG). This component generates power through the interaction of an electromagnetic field produced by rotor windings within an electromagnetic field produced by stator windings [100]. Power is generated at variable speed by a DFIG by varying the frequency of the electrical supply to stator windings independently of the rotor windings. The main components of a generator are the stator core and frame, rotor, and windings [100].

Faults within a generator can be electrical or mechanical [100] [92]. Electrical faults typically stem from the windings, and can include the breakdown of winding insulation due to contamination, cracking, stress or high temperatures [100]. This causes short circuits and electrical discharge within the windings which can lead to extreme high temperatures and unbalanced electromagnetic forces during rotation. Other electrical faults include currents induced within the stator core causing expansion and melting, or connection faults between brushes and the rotor [100]. Mechanical faults include misalignment, unbalance or looseness and can be caused by manufacturing imperfections, incorrect installation, high loading or unbalanced electromagnetic forces [92].

4.1.3.1 Generator Fault Characteristic Frequencies

Both mechanical and electrical faults can be diagnosed from analysis of vibration data within the frequency domain [92].

Mechanical faults occur due to faults within the shaft or supporting bearings, usually causing vibration components at the rotation frequency of the rotor with different harmonics [92].

Electrical faults can be due to faults within the generator's rotor or stator, introducing vibrations due to the interaction of unbalanced magnetic forces [92]. Such faults cause frequency components related to slip frequency (f_{slip}), pole-pass frequency (f_{p}) and rotor bar pass frequency ($RBPF$), calculated in equations (4.12) to (4.13) [92] where f_{L} is the line frequency, n_{poles} is the number of generator poles and n_{bars} is the number of bars in the rotor.

$$f_{\mathrm{slip}} = \frac{2 \times f_{\mathrm{L}}}{n_{\mathrm{poles}}} - f_{\mathrm{rotor}} \tag{4.12}$$

$$f_{\mathrm{p}} = f_{\mathrm{slip}} \times n_{\mathrm{poles}} \tag{4.13}$$

$$RBPF = n_{\mathrm{bars}} \times f_{\mathrm{rotor}} \tag{4.14}$$

4.1.3.2 Unbalance

Unbalance introduces a prominent spike at the rotation frequency of the rotor f_{rotor} which is always present and dominates the spectrum [92], as in Figure 4.15. The amplitude of this component varies proportionally to the square of the rotation speed.

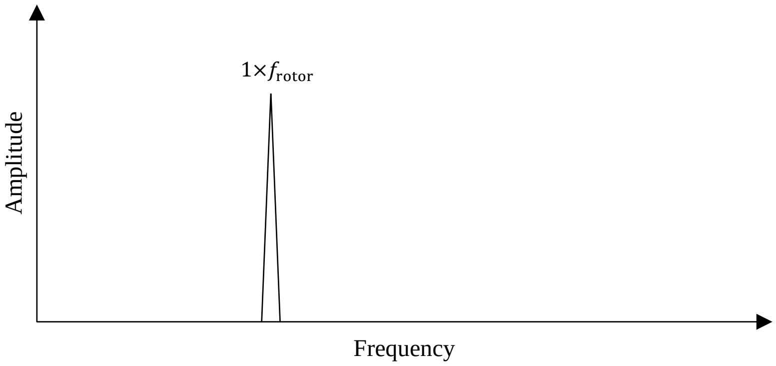


Figure 4.15 - Frequency response of unbalance within the generator, adapted from [92]

4.1.3.3 Bent Shaft

A bent generator shaft produces frequency components at 1x and 2x f_{rotor} [92]. The 1x component is dominant when the bend is near the centre of the shaft, as in Figure 4.16, and the 2x component is dominant when the bend is near the end of the shaft, as in Figure 4.17. This is more likely to produce vibrations across the axial direction as opposed to the radial direction.

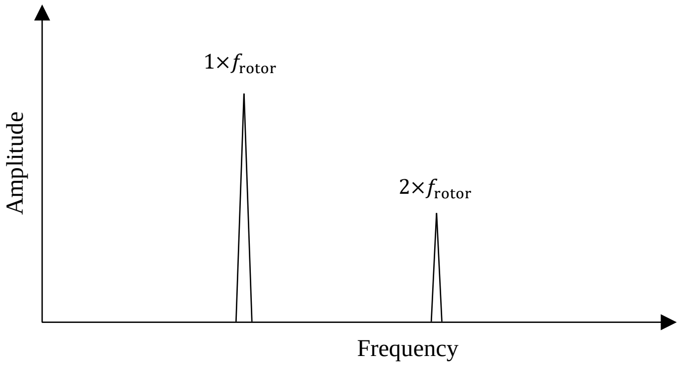


Figure 4.16 - Frequency response a bend near the centre of the generator shaft, adapted from [92]

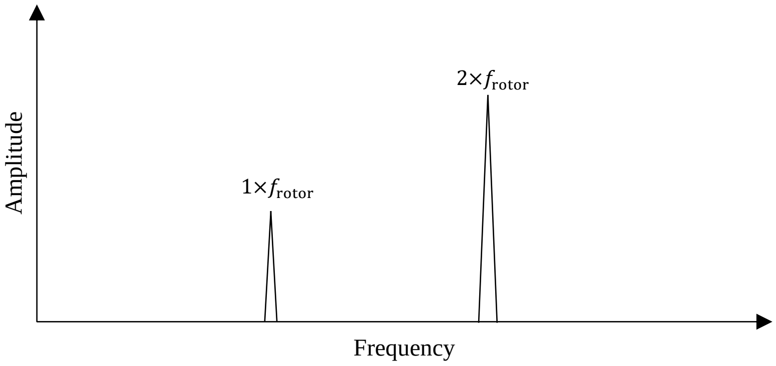


Figure 4.17 - Frequency response a bend near the centre of the generator shaft, adapted from [92]

4.1.3.4 Misalignment

Misalignment can be either angular or parallel [92]. Angular misalignment produces high axial vibrations at 1x, 2x and 3x f_{rotor}, Figure 4.18. Parallel misalignment also causes harmonics of f_{rotor} where the 2x component is most likely to dominate, Figure 4.19. As the severity of parallel misalignment increases, harmonics may appear at up to 8x f_{rotor}. Parallel misalignment, however, is unlikely to occur independently of angular misalignment.

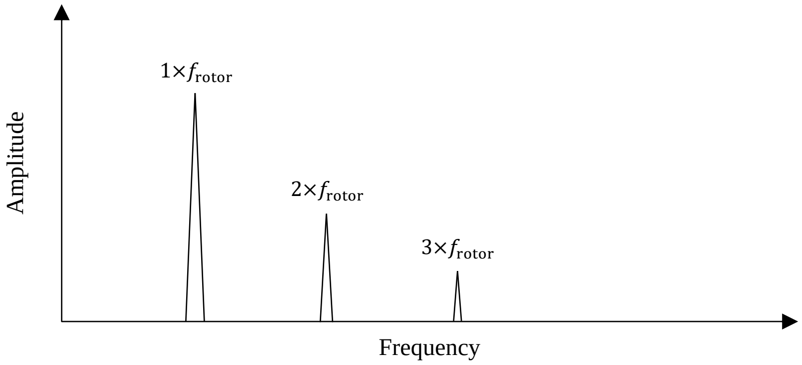


Figure 4.18 - Frequency response of angular generator misalignment, adapted from [92]

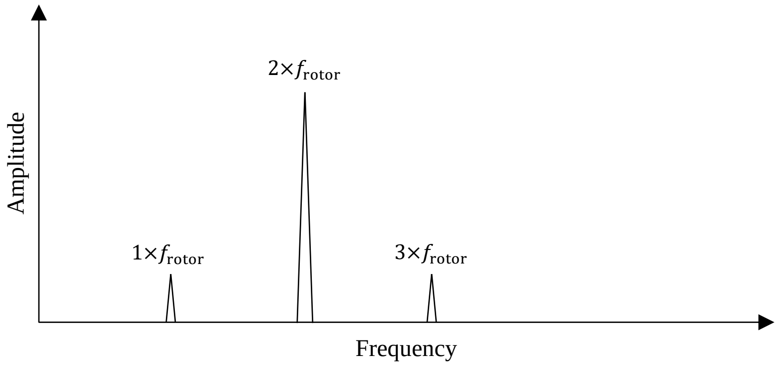


Figure 4.19 - Frequency response of parallel generator misalignment, adapted from [92]

4.1.3.5 Mechanical Looseness

Mechanical looseness can occur in three locations:

1. Internal assembly looseness
2. Looseness at machine to base plate interface
3. Structure looseness

Such looseness can produce many harmonics of f_{rotor} in the radial direction and can also induce sub-harmonic frequencies at $\frac{1}{2}$ or $\frac{1}{3}$ f_{rotor} [92], as seen in Figure 4.20.

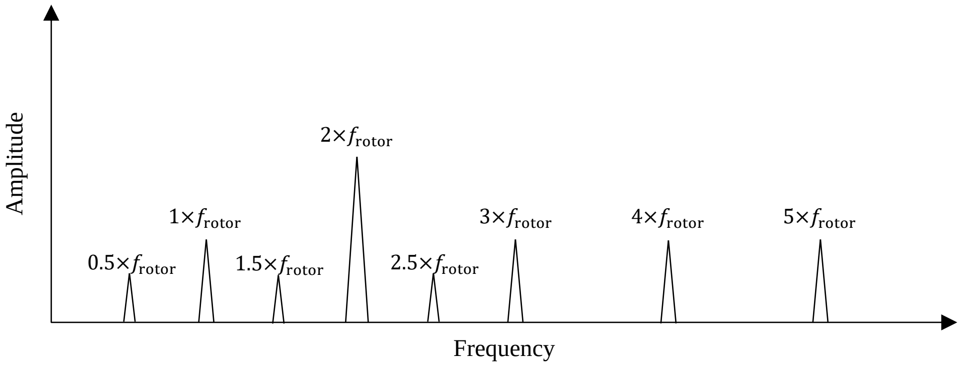


Figure 4.20 - Frequency response of generator mechanical looseness, adapted from [92]

4.1.3.6 Rotor Rubs

Rotor rubs produce a spectrum similar to that observed during mechanical looseness. A series of harmonics of f_{rotor} is generated, normally exciting integer fraction sub-harmonics $(\frac{1}{2}, \frac{1}{3}, \frac{1}{4} \dots \frac{1}{n})$ depending on the location of rotor natural frequencies [92].

Frequencies may occur as below, where N is the shaft speed and N_{c} is the critical speed of the shaft, a speed at which shaft deflections occur causing resonant vibrations.

$$\begin{array}{rl} 1\times f_{\mathrm{rotor}} & \text{when } N < N_{\mathrm{c}} \\ \frac{1}{2} \text{ or } 1\times f_{\mathrm{rotor}} & \text{when } N > 2N_{\mathrm{c}} \\ \frac{1}{3}\times, \frac{1}{2}\times \text{or } 1\times f_{\mathrm{rotor}} & \text{when } N > 3N_{\mathrm{c}} \\ \frac{1}{4}\times, \frac{1}{3}\times, \frac{1}{2}\times \text{ or } 1\times f_{\mathrm{rotor}} & \text{when } N > 4N_{\mathrm{c}} \end{array} \tag{4.15}$$

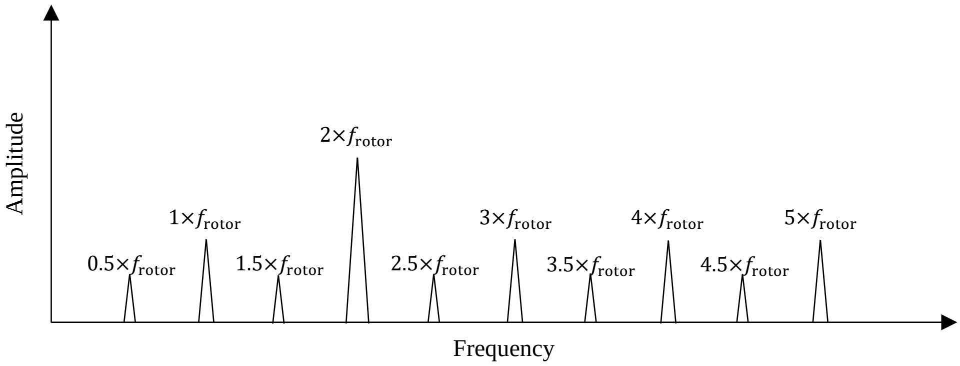


Figure 4.21 - Frequency response for rotor rubs in generator, adapted from [92]

4.1.3.7 Rotor Defects

Rotor defects causing unbalance of magnetic force between the rotor and stator will cause vibration in the radial direction. Broken or cracked rotor bars or shorting rings, bad joints between rotor bars and shorting rings or shorted rotor laminations will produce frequency components at f_{rotor} with sidebands of 2x f_{slip} [92], Figure 4.22.

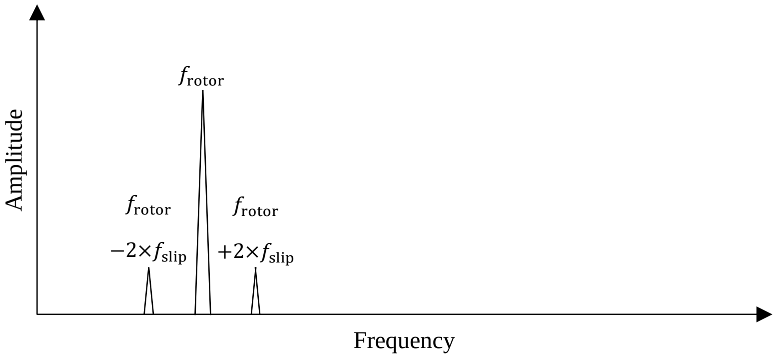


Figure 4.22 - Frequency response for rotor defects in the generator (e.g. breaks/cracks or bad joints in rotor bars or shorting rings, or shorted rotor laminations), adapted from [92]

In addition, cracked rotor bars often generate f_{p} sidebands around 3rd, 4th and 5th f_{rotor} harmonics [92], Figure 4.23.

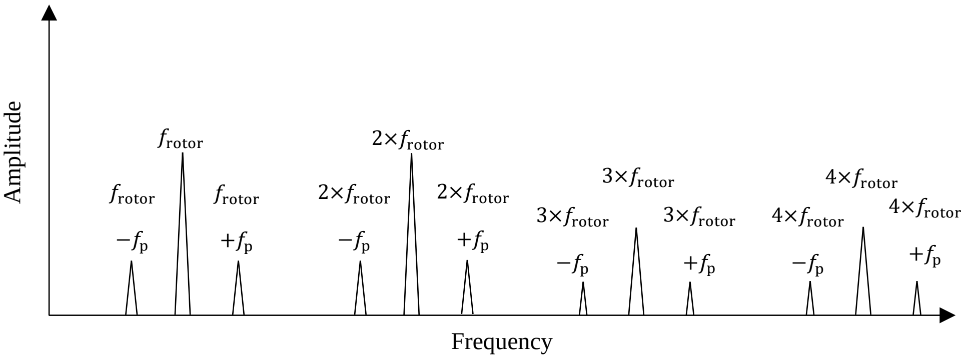


Figure 4.23 - Frequency response for cracked rotor bars in the generator, adapted from [92]

Finally, loose rotor bars are indicated by RBPF and sidebands of 2x line frequency [92], Figure 4.24.

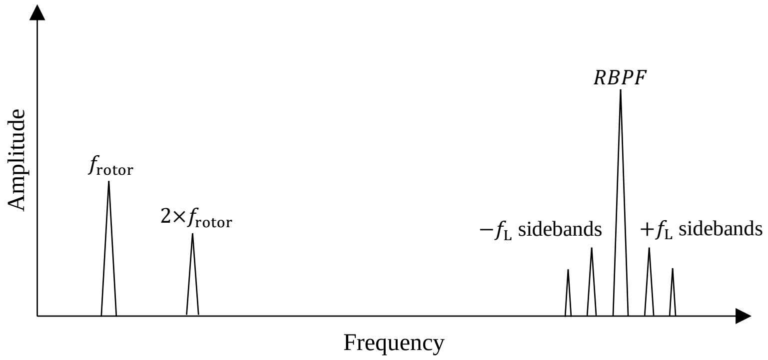


Figure 4.24 - Frequency response for loose rotor bars in the generator, adapted from [92]

4.1.3.8 Eccentric Rotor

Increasing eccentricity in the rotor can cause highly unbalanced magnetic forces. In the frequency spectrum, eccentric rotors will produce vibration components of 2x f_{L} surrounded by pole pass frequency sidebands (as well as f_{p} sidebands around 1x f_{rotor}) [92]. The pole pass frequency f_{p} may also appear at low frequency, Figure 4.25.

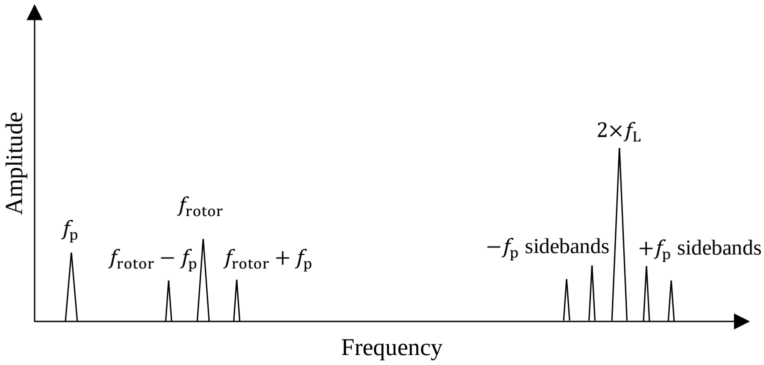


Figure 4.25 - Frequency response of eccentric rotor in the generator, adapted from [92]

4.1.3.9 Stator Defects

Stator problems or defects produce high vibration at 2x line frequency f_{L} [92], Figure 4.26. Eccentricity produces uneven stationary air gaps between the rotor and stator and can produce strong vibrations across single directions (i.e. across either the X, Y, or Z axis).

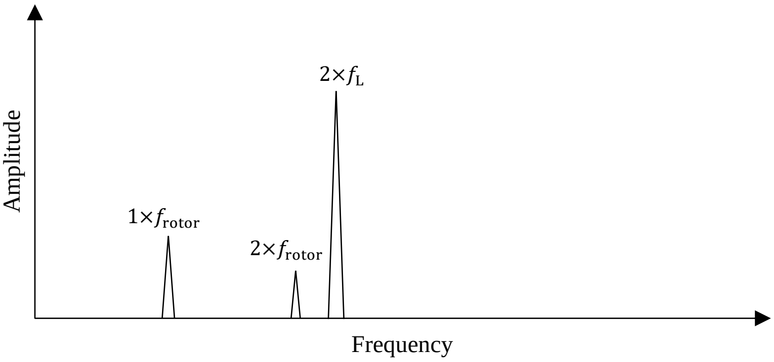


Figure 4.26 - Frequency response of stator defects, adapted from [92]

4.1.3.10 Phasing Problem

Phasing problems due to loose or broken connectors can cause excessive vibration at 2x line frequency f_L, which have sidebands at 1/3 of the line frequency [92], Figure 4.27.

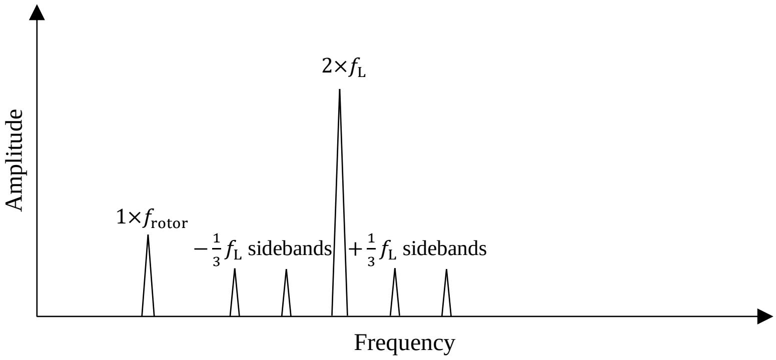


Figure 4.27 - Frequency response of phasing problems within the generator, adapted from [92]

4.2 Feature-Based Diagnostics

Following anomaly detection, a fault diagnosis can be manually made by engineers inspecting vibration data to identify the characteristics detailed in section 4.1. However, the aim of intelligent condition monitoring systems is to provide automatic fault detection and diagnosis without manual inspection of sensor data.

Methods of automatic fault diagnosis are commonly performed through feature-based approaches [29]. In this process, a number of signal processing stages are used to transform vibration signals into feature vectors, isolating and quantifying specific characteristics (known as features) from within vibration data representative of different fault modes. This is known as feature-extraction. The resultant feature data can then be automatically categorised using machine learning classification algorithms. This process is summarised in Figure 4.28.

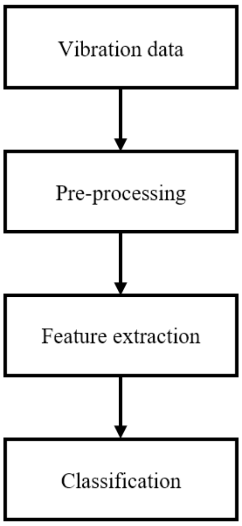


Figure 4.28 - High level feature-extraction process for rotating machine diagnostics from vibration data

4.2.1 Pre-processing

A pre-processing stage may be required in some applications to modify raw vibration signals before feature-extraction. Two such techniques are presented here: order tracking (for normalising vibration signals recorded under varying rotation speeds) and enveloping (required in the detection of bearing faults).

4.2.1.1 Order Tracking

The frequency values of many fault characteristics discussed in section 4.1. are dependent on the rotation speed of the machinery. In practical applications, and in particular within tidal power, the rotation speed is not constant and can fluctuate rapidly over short periods of time. This can therefore obscure vibration signals in the frequency domain, as individual vibration components are spread over a range of frequency values as the rotation speed of the machine changes over time.

A solution is to utilise order tracking techniques as a pre-processing stage before signals are analysed in the frequency domain. Order tracking involves the transformation of a signal from the time domain, into the order (or angular) domain, where vibrations are represented in terms of multiples of shaft rotation [101]. Two order tracking techniques are discussed in this work: computed resampling and the Vold-Kalman filter.

4.2.1.1.1 Computed Resampling

Computed order tracking involves resampling vibration data from being sampled at fixed intervals in time $x(t)$ to fixed intervals of the rotation of a shaft $x(\theta)$ [101]. This process is detailed in Figure 4.29.

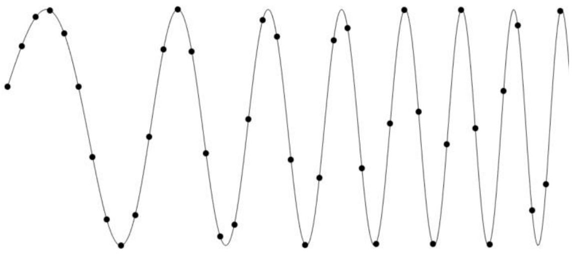
The resampled signal $x(\theta)$ can be calculated using interpolation, where signal $x(t)$, sampled at points $t(i\Delta t)$ over time, is interpolated at sample points $t(i\Delta\theta)$ over time. $\Delta\theta$ is calculated as in (4.16) where N is an integer multiple of shaft rotation θ and f_r is the rotation frequency of the shaft at point i.

$$\Delta\theta = \frac{1}{N f_r(i)} \tag{4.16}$$

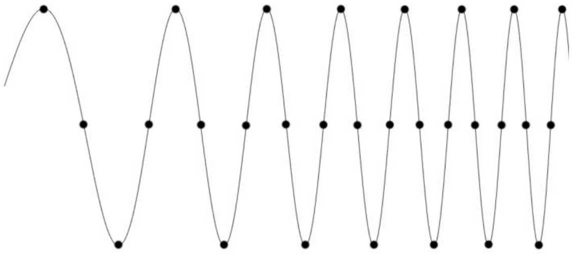
Performing a frequency domain transformation on $x(\theta)$, such as the FFT, will express the signal in the order domain, as opposed to the frequency domain, where the frequency axis is replaced by an order axis. Values on the order axis $\phi(i)$ represent multiples of shaft revolutions, described in (4.17), where $f(i)$ are the frequency axis values in Hz and f_r is the shaft rotation frequency.

$$\phi(i) = \frac{f(i)}{f_r} \tag{4.17}$$

This results in responses as illustrated in Figure 4.30. Vibration components whose frequency values are dependent on rotation speed are represented more distinctly on the order axis than the frequency axis.

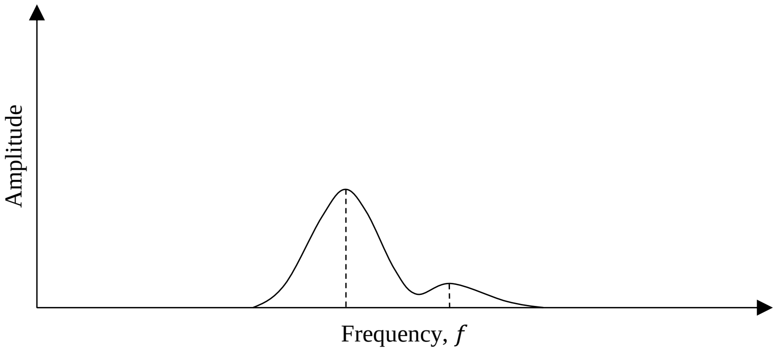


(a) Varying frequency signal $x(t)$ sampled at regular time intervals $t(i)$

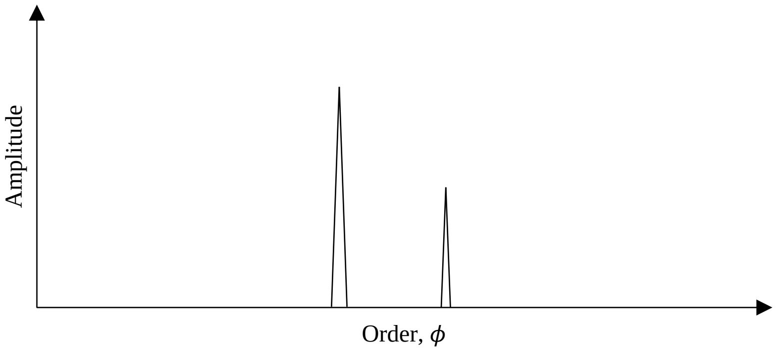


(b) Varying frequency signal resampled at regular angular rotation intervals of $\theta(i)$ where $\Delta\theta = \frac{1}{4}$

Figure 4.29 - Computed resampling applied to vibration signal during variable rotation speed



(a) Vibration signal in the frequency domain during variation in rotation speed



(b) Vibration signal in the order domain during variation in rotation speed

Figure 4.30 - Spectrum of vibration signal before and after resampling

4.2.1.1.2 Vold-Kalman Filter

An alternative order tracking technique is the Vold-Kalman filter. This technique acts as a moving bandpass filter, set to the frequency values of specific vibration components over varying rotation speeds. Given an RPM measurement and set of orders to be extracted, this technique consists of solving a set of linear least squares equations known as the structural and data equations to form a filtered signal [102].

The structural equation (4.18) details a second-order difference equation [102], where $x(n)$ is the desired frequency component to be filtered, a sine wave with frequency ω sampled with a period of Δt at points n.

$$x(n\Delta t) - 2\cos(2\pi\omega\Delta t)\, x\big((n-1)\Delta t\big) + x\big((n-2)\Delta t\big) = 0 \tag{4.18}$$

This can be rewritten as (4.19), where Δt is dropped, $c(n) = \cos(2\pi\Delta\omega)$ and term $\varepsilon(n)$ is introduced, allowing $x(n)$ to change its amplitude, phase and frequency over time.

$$x(n) - c(n)x(n-1) + x(n-2) = \varepsilon(n) \tag{4.19}$$

Equation (4.20) details the data equation [102], where measured signal $y(n)$ is the composition of the desired vibration feature $x(n)$ that satisfies the structural equation and $\eta(n)$ representing noise and other periodic components.

$$y(n) = x(n) + \eta(n) \tag{4.20}$$

Equations (4.19) and (4.20) can be rearranged to give equation (4.21) to provide a set of linear equations for $\{x(n) \quad x(n-1) \quad x(n-2)\}$.

$$\begin{Bmatrix} 1 & -c(n) & 1 \\ & & 1 \end{Bmatrix} \begin{bmatrix} x(n-2) \\ x(n-1) \\ x(n) \end{bmatrix} = \begin{bmatrix} \varepsilon(n) \\ y(n) - \eta(n) \end{bmatrix} \tag{4.21}$$

A weighting term $r(n)$, detailed in equation (4.22), of the ratio of standard deviations of $\varepsilon(n)$ and $\eta(n)$ is then applied as in equation (4.23).

$$r(n) = \frac{s_{\varepsilon}(n)}{s_{\eta}(n)} \tag{4.22}$$

$$\begin{Bmatrix} 1 & -c(n) & 1 \\ & & r(n) \end{Bmatrix} \begin{bmatrix} x(n-2) \\ x(n-1) \\ x(n) \end{bmatrix} = \begin{bmatrix} \varepsilon(n) \\ r(n)\big(y(n) - \eta(n)\big) \end{bmatrix} \tag{4.23}$$

Equation (4.23) can then be solved for the desired filtered signal $x(n)$ using least-squares techniques [102].

4.2.1.2 Enveloping

Certain faults within rotation equipment (primarily within bearings) cause periodic impacts at natural frequencies which can be seen within vibration signals, as described in section 4.1.2.1, p. 90. The periodic component of these fault signals (required to make a diagnosis) may be difficult to observe within the time and frequency spectrum, where natural frequencies can be dominant. Enveloping can be used as a pre-processing technique to reveal the periodic component of vibration signatures caused by periodic impact faults before frequency analysis is performed.

This enveloping process is detailed in Figure 4.31 for a simulated bearing fault. In this process, a bandpass filter is first applied at the natural frequency (commonly between 3 and 50 kHz for bearings [99]). The envelope of this filtered signal can then be found using the Hilbert transform.

The Hilbert transform computes the analytic signal $z(t)$ from time-domain signals $x(t)$. The analytic signal represents the Fourier transform of a signal, but with no negative frequency components. This can be calculated for discrete-time signals as follows [103]:

1. Calculate N-point Discrete-Time Fourier Transform (DTFT) (4.24) of filtered time-domain vibration signal $x_{\text{filt}}(t)$, where f is a set of discrete frequency values and T is the sampling period.

$$X(f) = T\sum_{n=0}^{N-1} x_{\text{filt}}(t)e^{(-j2\pi f tT)} \tag{4.24}$$

2. Compute an N-point analytic signal transform $Z(f)$ as in (4.25).

$$Z(f) = \begin{cases} X(0), & \text{for } m = 0 \\ 2X(f), & \text{for } 1 \leq f \leq \dfrac{N}{2} - 1 \\ X\left(\dfrac{N}{2}\right), & \text{for } f = \dfrac{N}{2} \\ 0, & \text{for } \dfrac{N}{2} + 1 \leq m \leq N - 1 \end{cases} \tag{4.25}$$

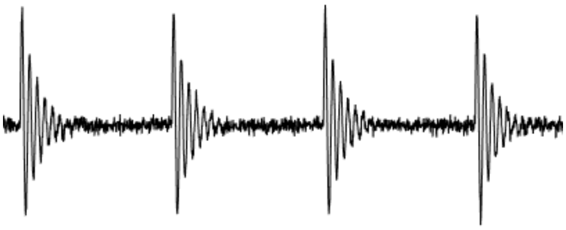
3. Use the inverse DTFT (4.26) to compute analytic signal $z(t)$.

$$z(t) = \frac{1}{NT}\sum_{m=0}^{N-1} Z(f)e^{\left(\frac{j2\pi f t}{N}\right)} \tag{4.26}$$

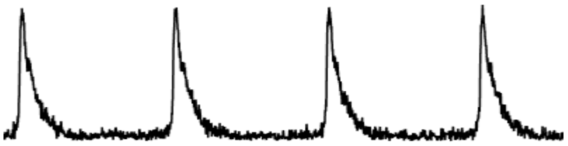
4. The envelope $e(t)$ of the filtered vibration signal $x_{\text{filt}}(t)$ can then be calculated using the analytic signal $z(t)$ as in (4.27).

$$e(t) = \sqrt{x_{\text{filt}}(t)^2 + z(t)^2} \tag{4.27}$$

(a) Vibration signal $x(t)$



(b) Vibration signal bandpass filtered at natural frequency $x_{filt}(t)$



(c) Envelope signals $e(t)$

Figure 4.31 - Envelope detection process for detecting periodic impacts in a simulated vibration signal for a bearing fault

4.2.2 Feature-extraction

Feature-extraction is then performed following pre-processing (Figure 4.28, p. 100). Here, fault characteristics are isolated from processed vibration signals, transforming the original data into a vector of features representative of individual fault modes [29].

Effective signal processing techniques for isolating and quantifying rotating machine faults include time, frequency and time-frequency domain transformations [29].

4.2.2.1 Time Domain

A number of different metrics can be used to define characteristics of vibration sensor data in the time domain.

Root mean square (RMS) is a time-domain metric used to measure the power content of a signal [104]. This metric is useful in describing the overall noise level produced by a vibrating component, where faults will introduce excessive noise to the vibration signal. RMS can be calculated for a time domain signal x over a set window size N as in (4.28).

$$x_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{4.28}$$

The crest factor, the ratio between peak and RMS values of vibration [104], is calculated as in (4.29). This parameter gives an indication of how ‘impulsive’ data is, i.e. contains large isolated peaks. Certain faults within rotating equipment will introduce periodic impacts (e.g. a cracked gear tooth or bearing ring fault), increasing the impulsiveness of a vibration signal, consequently increasing the crest factor, allowing this metric to be used for fault identification and diagnosis.

$$C = \frac{|x|_{\mathrm{peak}}}{x_{\mathrm{RMS}}} \tag{4.29}$$

Kurtosis is a measure of the spread of data from the mean value, describing if the distribution is peaked or flat [104]. Damage to rotating components will introduce additional vibration peaks, therefore increasing the kurtosis of the signal. Kurtosis is calculated as shown in equation (4.30), where $\bar{x}$ and σ^2 are the mean and variance of the signal respectively over window size N.

$$K = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{(N\sigma^2)^2} \tag{4.30}$$

Skewness is a measure of the symmetry of a distribution [105]. This can be a useful metric in describing vibration where faults can cause the distribution of vibration data to tend towards more positive values. This can be calculated as in (4.31).

$$K = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{N\sigma^3} \tag{4.31}$$

FM0 is a metric used to describe gear faults by detecting changes in gear mesh patterns [106]. It is calculated as in (4.32), where $x_{\mathrm{p-p}}$ is the peak to peak amplitude of vibration signal x and $A(f_i)$ is the amplitude of N gear mesh frequencies and their harmonics.

$$FM0 = \frac{x_{\mathrm{p-p}}}{\sum_{i=1}^{N} A(f_i)} \tag{4.32}$$

FM4 is a metric used to complement FM0 by describing gear faults limited to a set number of gear teeth [106]. This is calculated as in (4.33), equivalent to the normalised Kurtosis of d, a difference signal of length N where first order sidebands have been removed from the raw vibration.

$$FM4 = \frac{N \sum_{i=1}^{N} \left(d_i - \bar{d}\right)^4}{\left[\sum_{i=1}^{N} \left(d_i - \bar{d}\right)^2\right]^2} \tag{4.33}$$

NA4 is a metric developed to detect and measure gear damage in a vibration signal as it grows and increases in amplitude [107]. This is calculated as in (4.34) where r is a residual signal of length N where fundamental shaft and mesh frequencies (and harmonics) have been removed from the raw vibration signal after m points in time.

$$NA4 = \frac{N \sum_{i=1}^{N} (r_i - \bar{r})^4}{\left[\frac{1}{m} \sum_{j=1}^{m} \left[\sum_{i=1}^{N} \left(r_{ij} - \bar{r}_j\right)^2\right]\right]^2} \tag{4.34}$$

NA4* is an enhanced version of NA4, more robust to changes in variance of the signal as damage grows [108]. This is calculated as in (4.35) where $\widetilde{M}$ is the variance of residual signal r for the gearbox in a healthy condition.

$$NA4^* = \frac{\frac{1}{N} \sum_{i=1}^{N} (r_i - \bar{r})^4}{\left(\widetilde{M}_2\right)^2} \tag{4.35}$$

NB4 is similar to NA4, however uses the envelope signal E (explained in section 4.2.1.2, p. 104) to measure damage in machinery [107] and is calculated as in (4.36).

$$NB4 = \frac{N \sum_{i=1}^{N} (E_i - \bar{E})^4}{\left[\frac{1}{m} \sum_{j=1}^{m} \left[\sum_{i=1}^{N} \left(E_{ij} - \bar{E}_j\right)^2\right]\right]^2} \tag{4.36}$$

M6A is an indicator of surface damage on machinery [109]. This metric is based on FM4, however uses the 6th moment to be more sensitive to peaks within the difference signal d, calculated as in (4.37).

$$M6A = \frac{N^2 \sum_{i=1}^{N} \left(d_i - \bar{d}\right)^6}{\left[\sum_{i=1}^{N} \left(d_i - \bar{d}\right)^2\right]^3} \tag{4.37}$$

M8A is an extension of M6A to detect surface damage of machinery in the difference signal d, and is calculated as in (4.38), utilising the 8th moment normalised by variance to the 4th power [109].

$$M8A = \frac{N^3 \sum_{i=1}^{N} \left(d_i - \bar{d}\right)^8}{\left[\sum_{i=1}^{N} \left(d_i - \bar{d}\right)^2\right]^4} \tag{4.38}$$

4.2.2.2 Frequency Domain

Although time domain analysis of vibration signals can give useful indicators of rotating machinery faults, features extracted in the frequency domain are more descriptive of characteristic vibration signals [110] (as described in section 4.1, p. 83) and can therefore differentiate between specific fault modes more effectively.

Frequency domain vibration analysis is widely performed using the Fast Fourier Transform (FFT) [111]. The FFT is an implementation of the Discrete Fourier Transform (DFT), where time domain signals $x(t)$ are transformed and represented within the frequency domain $X(f)$ [112]. An N point DFT is described in equation (4.24). For example, N. Baydar, et al. [113], use the power spectrum (the square of the amplitude spectrum obtained through the FFT) to detect characteristic frequencies and locate gear tooth wear under varying loading conditions.

Frequency domain transformation can also be used with time-domain vibration metrics to further analyse and quantify characteristic frequency components. P. Tamilselvan, et al [114], identify bearing faults and planetary gear mesh faults through the FFT for a wind turbine gearbox. A severity metric is then assigned to each frequency based on vibration kurtosis, a statistical measure of the peak of a probability distribution.

However, frequency domain analysis is mostly only suitable for stationary vibration signals, where the distribution of vibration signals does not change over time [111], for example for fault signals such as misalignment and unbalance.

4.2.2.3 Time-frequency domain

While traditional frequency domain transforms, such as the FFT, assume signals are periodic and stationary [111], rotating machine faults can often be non-stationary, where the distribution of a vibration signal is not constant over periods of time. These signals can occur due to contact between cracked or broken gear teeth, or due to periodic impacts caused during bearing faults [115].

More modern feature-extraction methods therefore use time-frequency methods of feature-extraction can be useful for isolating such non-stationary fault signals, where the change in frequency response is analysed over time [110]. Methods include the Short Time Fourier Transform and the Wavelet Transform.

4.2.2.3.1 Short-Time Fourier Transform

The Short-Time Fourier Transform (STFT) is a Fourier transform performed on long periods of data repeatedly over short periods of time [116]. This gives a representation of the change in frequency response over time, allowing non-stationary effects within signals to be observed.

The STFT is typically performed by applying a sliding windowing function to data and applying the Fourier transform as the window is shifted along the time axis. The continuous STFT for signal $x(t)$ is given in (4.39) [116], where $w(t-\tau)$ is a window function (e.g. Hamming window) shifted by τ.

$$X(f,t) = \int_{-\infty}^{\infty} x(\tau)w(t-\tau)e^{j2\pi f\tau}\,d\tau \tag{4.39}$$

The continuous STFT can be modified to the discrete STFT as in (4.40) [116], where $x(k)$ is the sampled signal, T is the sampling period, $w(nD-k)$ is a shifted window function at sampling point n with period D.

$$X(n,m) = \sum_{k=0}^{T-1} x(k)w(nD-k)e^{\frac{j2\pi km}{T}} \tag{4.40}$$

A spectrogram, the visualisation of the frequency response over time, can be formed by taking the magnitude of the discrete STFT. Figure 4.32 shows an example of a spectrogram for simulated vibration signal for a bearing fault, highlighting non-stationary signals, where time is on the X-axis, frequency on the Y-axis and magnitude reflected through colour scale.

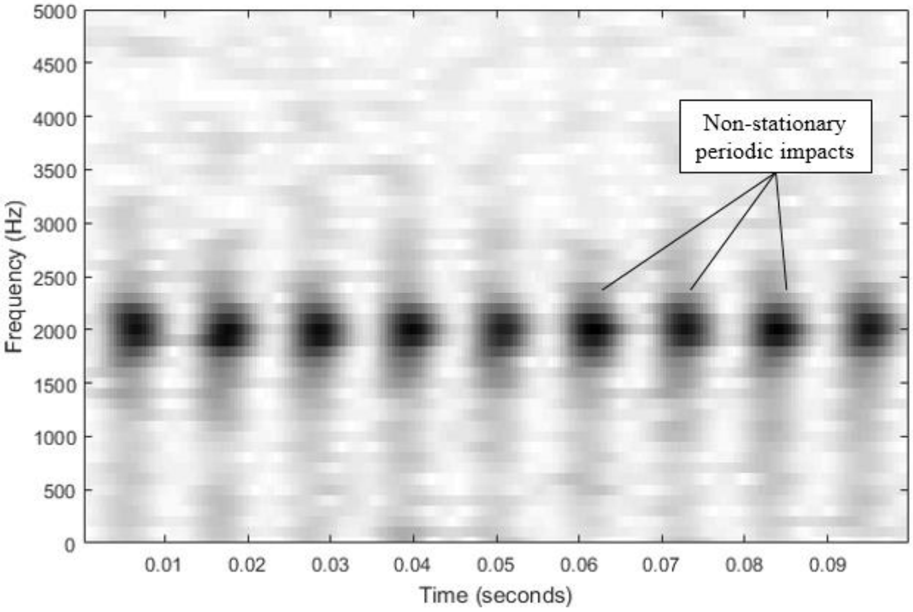


Figure 4.32 - Spectrogram of simulated bearing fault generated through the STFT, highlighting non-stationary vibration produced through periodic impacts

M-C. Pan, et al. [117] show how the STFT provided improved clarity in comparison to time and frequency domain analysis for non-stationary signals caused by gear transmission noise and brake systems. N. Sawalhi and R. B. Randall [118] show how the STFT, coupled with a computed resampling order tracking algorithm, can be used to detect gear mesh frequencies during periods of varying rotation speed. Finally, M. S. Safizadeh, et al. summarise the use of the STFT in machine diagnostics [119]. It is stated that while the STFT gives a clear representation of the time-frequency domain that allows detection of faults, a major shortcoming of the technique is its inability to have high resolution in both the time and frequency axes simultaneously. As a result, some time may have to be spent tuning the size of the sliding window (as is (4.40)) to give clear results depending on the length of available data and the desired frequency range.

4.2.2.3.2 Wavelet Transform

The Wavelet Transform is an alternative time-frequency transformation technique, and is becoming commonly used to identify non-stationary fault signals within vibration data [91].

Within the continuous wavelet transform a mother wavelet function, scaled in frequency and shifted in time, is convolved with a time domain signal, producing a time-frequency representation of the signal [91]. Equation (4.41) details this process, where a is a scaling factor, b is a time-shift parameter and Ψ is the mother wavelet function.

$$W(a;b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\Psi^{*}\left(\frac{t-b}{a}\right) dt \tag{4.41}$$

Figure 4.33 shows a visualisation of the continuous wavelet transform for a simulated periodic impact bearing fault, where scaling factor a and time-shift parameter b have been converted to represent frequency and time values respectively.

The wavelet transform can be applied to vibration signals to diagnose faults, most commonly within gearboxes and bearings to detect non-stationary fault signals [120] [121] [122] [123] [124]. For example, M. Amarnath, et al [121], use the continuous wavelet transform applied to vibration signals to identify and diagnose the extent of a gear tooth crack within a two-stage helical gearbox test rig. J. Rafiee, et al [123], compare a number of mother wavelet functions to extract features from vibration data to diagnose gear and bearing faults, such as ball faults, cage faults, inner race faults in bearings and broken teeth in gearboxes.

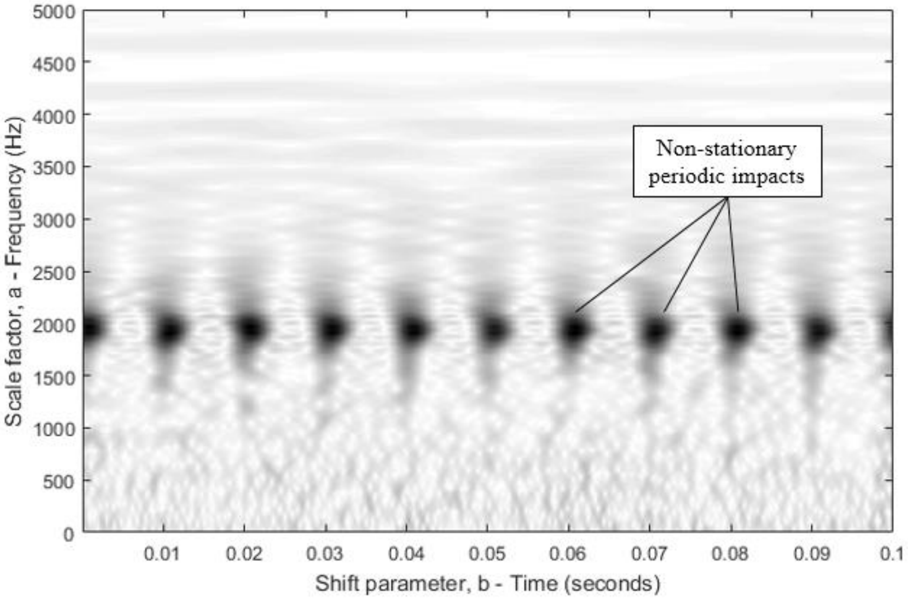


Figure 4.33 - Visualisation of the Continuous Wavelet Transform for a simulated bearing fault, highlighting non-stationary vibration produced through periodic impacts

4.2.2.4 Cepstrum Analysis

Characteristic frequencies of rotating components often contain sidebands of harmonics in addition to core frequency values. This can make interpretation of vibration through the frequency spectrum (obtained through an FFT or otherwise) challenging, as multiple families of sidebands and harmonics obscure the spectrum.

Cepstrum analysis provides a solution, where families of related frequency components from a single characteristic frequency are represented as single components. This is obtained through the inverse Fourier transform of the logarithm of the power spectrum (4.42) [125], transforming the signal into the 'quefrency' domain τ where F_{xx} is the power spectrum of signal $f_x(t)$ (4.43).

$$C(\tau) = \mathcal{F}^{-1}\{\log F_{xx}(f)\} \tag{4.42}$$

$$F_{xx}(f) = |\mathcal{F}\{f_x(t)\}|^2 \tag{4.43}$$

The quefrency domain is measured in time, where amplitudes within this domain represent the spacing of sidebands within the spectrum at given periods. Large amplitudes represent fast variations in the spectrum, relative to a large number of sidebands, and small amplitudes represent slow variations in the spectrum, relative to a low number of sidebands [125].

Within diagnosis, faults introducing additional sidebands to the vibration spectrum can be identified as peaks within the Cepstrum. The period of these peaks can be used for diagnosis are they are relative to the fault's characteristic frequency. Figure 4.34 demonstrates the quefrency spectrum for gearbox vibration data from a worn tooth. Peaks spaced at T_{gear} are observed on this spectrum, the period of the gear rotation speed.

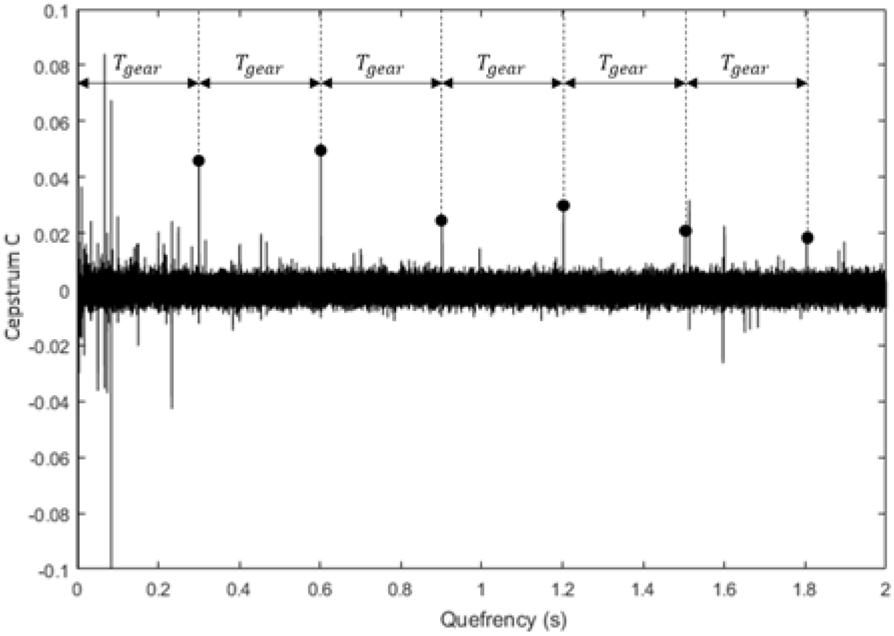


Figure 4.34 - Quefrency domain signal for a cracked gear tooth

M. Satyam, et al [126] use Cepstrum analysis applied to vibration signals from a wind turbine gearbox test rig, identifying quefrency peaks relating to gear shaft misalignment. They state this method is suitable for diagnosing other faults within gears, bearings and turbine blades, where families of harmonics and sidebands exist.

L. Nacib, et al [127], use Cepstrum analysis to identify a tooth crack within a helicopter gearbox. Periodic peaks within the Cepstrum are identified and used to diagnose a cracked tooth within a particular gear pinion.

4.2.3 Classification

The final stage of a feature-based diagnostics system is the classification stage (Figure 4.28, p. 100), where faults are automatically diagnosed from selected features.

This can be performed through a machine learning classification algorithm trained through supervised learning, whose inputs are features and outputs are discrete fault states. Three different classification techniques were examined in this study: support vector machines, decision trees and k-nearest neighbours.

4.2.3.1 Support vector machines

Support vector machines (SVMs) have been discussed in chapter 3, section 3.5.5.1, p. 54, for one-class anomaly detection, however they are more commonly used as a classification technique.

SVMs are used to define a hyperplane representing boundaries in multidimensional space between different classes of data [82]. In rotating machine diagnostics, each dimension in the data space may represent a different feature extracted from vibration signals, where each class represents a different health state or fault mode.

In the simplest cases, hyperplanes are defined linearly. Linear hyperplanes can either be defined through a hard-margin or soft-margin, where the margin is defined by the hyperplane and the most adjacent data points on either side [128]. A hard-margin is used when classes can be separated linearly in the data space and a soft-margin is used when classes are not linearly separable.

Given training data $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)$ where y_i values are class labels of 1 or -1, data is linearly separable when (4.44) and (4.45) are satisfied for vector $\boldsymbol{w}$ and scalar b [82].

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b \geq 1 \quad \text{for } y_i = 1 \tag{4.44}$$

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b \leq -1 \quad \text{for } y_i = -1 \tag{4.45}$$

Expressions (4.44) and (4.45) can be rewritten as in (4.46).

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq \ +1 \tag{4.46}$$

Therefore, in the hard-margin case, a linear hyperplane maximising the distance between data points $\boldsymbol{x}_i$ and the hyperplane in each class can be found by finding $\boldsymbol{w}$ and b that satisfy (4.47).

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0 \tag{4.47}$$

However, in many cases data are not linearly separable. In such cases, a soft-margin hyperplane is defined, allowing some data points to lie on the wrong side of each margin, but with the least error. Variables ξ_i can be introduced, storing the distance of each data point from the margin. If $\xi_i = 0$, the data lies on the correct side of the margin. If $0 < \xi_i < 1$, the data is correctly classified but lies within the margin. If $\xi_i > 0$ the data point lies on the wrong side of the hyperplane and is misclassified [128]. These variables can therefore be added to (4.46) as in (4.48).

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1 - \xi_i \tag{4.48}$$

An optimal solution is found by finding $\boldsymbol{w}$ to minimise a cost function defined in (4.49), where C is a regularisation penalty factor [128].

$$\frac{1}{n} C \sum_{i=1}^{n} \xi_i + \frac{1}{2} \|\boldsymbol{w}\|^2 \tag{4.49}$$

Support vector machines are linear in nature. However, they can be adapted to solve non-linear problems by applying the ‘kernel trick’ [129]. As opposed to fitting a non-linear model, the kernel trick involves mapping the data into a new dimensional space through a nonlinear transformation and applying a linear model in the new space [130].

This transformation is described as in (4.50), where $\boldsymbol{x}$ is the original data space, $\boldsymbol{z}$ is the new dimension space and ϕ is a nonlinear kernel function [128].

$$\boldsymbol{z} = \phi(\boldsymbol{x}) \tag{4.50}$$

The hyperplane is now altered from (4.48) to become (4.51), where the linear hyperplane is defined in the new dimensional space $\boldsymbol{z}$ [128]. The hyperplane is then fitted in the new space by minimising the same cost function as in (4.49). This optimal linear hyperplane in the new space $\boldsymbol{z}$ is nonlinear in the original data space $\boldsymbol{x}$.

$$y_i(\boldsymbol{w} \cdot \phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i \tag{4.51}$$

SVMs are widely used in feature-based fault diagnosis systems [131] [132] [133] [134] [135] [136] [137] [138]. For example, Elangovan, et al. [131] compare SVMs of different kernel functions to other classification techniques (including decision trees) in diagnosing faults from statistical vibration features, finding SVMs with polynomial functions to provide the best performance. However, linear SVMs were found to perform classification at a faster rate, but at the sacrifice of some classification accuracy. P Konar, et al. [132] use an SVM classifier to diagnose bearing faults from features generated through the continuous wavelet transform, finding the technique to provide fast and efficient results. V. Sugumaran, et al. [133] compare the accuracy of SVM classifiers trained with different numbers of statistical input features (e.g. RMS, kurtosis, skewness, etc.). Optimal performance was achieved where 7 different features were used as inputs, however performance decreased using additional feature, showing SVMs are not robust to redundant features.

In general, SVMs give good classification performance when dealing with continuous features in multiple dimensions [139], and are therefore a good candidate for fault classification from vibration data features. Complex models can take some time to train, however, once models are trained data can be classified at high speeds [140], making SVMs suitable for use in online diagnostics systems.

Nevertheless, the performance of SVMs is highly dependent on the selection of a suitable kernel function for non-linear models. In addition, complex models can be hard to interpret, and limited engineering knowledge can be inferred from a trained model [140]. Finally, SVMs are natively binary classifiers [139], so multiple models may have to be constructed to detect multiple fault modes.

4.2.3.2 Decision tree

A decision tree is a structure containing a number of nodes that define a set of decision rules, predicting a set of outcomes (or class values for classification) from input data [141]. Internal nodes have 'branches' leading to subsequent nodes, performing decisions on the input data. 'Leaves' are nodes with no further decision branches and represent the most appropriate class for the input data leading to that point.

Decisions at each node make comparisons on only one dimension in $\boldsymbol{x}$ and can take two forms [128]. If $\boldsymbol{x}_i$ contains discrete values, i.e. describes categorical data of n possible values, the decision makes n splits, comparing the data to each of its possible values. If $\boldsymbol{x}_i$ contains numerical data, decisions take the form of (4.52), where c is any real number.

$$x_i \geq c \tag{4.52}$$

Classification trees can be constructed from top to bottom, aiming to maximise decisions leading to only one class (known as pure nodes). The probability of class $y = C_j$ at node m can be estimated as in (4.53) [128], where N_m is the number of training examples reaching node m. A pure node therefore has p_m^j for all J classes is equal to 0 or 1.

$$\hat{P}\left(C_j \middle| \boldsymbol{x}, m\right) \equiv p_m^j = \frac{N_m^j}{N_m} \tag{4.53}$$

Impurity functions are used to measure the effectiveness of splits at each node. One example of such a function is the Gini index [128], detailed in (4.54). Pure nodes have a Gini index of 0, otherwise the Gini index is a positive value.

$$I_G(p) = \sum_{j=1}^{J} p_j(1 - p_j) \tag{5.54}$$

The Classification and Regression Tree (CART) algorithm [142] uses the Gini index to construct decision trees. In this algorithm, the Gini index is calculated for all possible splits at a node, choosing the decision returning the lowest Gini index value. This is repeated for all subsequent nodes resultant from each split, until nodes become pure.

Decisions trees have been applied to the diagnosis of rotating machinery from vibration data in a number of cases [143] [144] [145] [146] [147] [148]. For example, N. T. Nguyen, et al. [143] use decision trees with both continuous and discrete features generated from vibration data to identify looseness, a stator fault, a rub fault and unbalanced rotor faults. H-H. Lee, et al. [144] apply decision trees to time domain vibration features for the diagnosis of bearing faults with high classification accuracy. B. S. Yang, et al. [145] apply decision trees to frequency domain vibration features for the diagnosis of multiple rotating machine faults (e.g. unbalance, misalignment, cracks, bearing damage, and many more) finding the method to perform well in creating interpretable cause-failure trees.

A primary advantage of decision trees over other classification techniques is their high interpretability [140], meaning the decision process can be easily understood by engineers and can be used to develop knowledge. As with SVMs, decision trees can classify data at high speeds and are therefore suitable for online monitoring systems [140]. Unlike SVMs however, decision trees are capable of classifying multiple classes through a single model [149] and can natively support both continuous and discrete inputs [140]. Decision trees can also have higher performance compared to other techniques when only small amounts of training data are available and are robust to irrelevant features [150]. However, decision trees can be prone to overfitting training data and can struggle with generalisation without the use of advanced ensemble methods [149].

4.2.3.3 K-nearest neighbours

K-nearest neighbours (K-NN) [151] is a non-parametric technique where classification is performed by measuring the distance between test data and training data. A class is assigned to test data by selecting the majority class of k training data points closest to the test data.

The Euclidean distance metric (4.55) is generally used where input data is numerical [151], where $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ are Cartesian coordinates of the test and training data points respectively in n-dimensional space.

$$d(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{4.55}$$

K-NN methods have been employed successfully in a number of cases for classifying vibration features for diagnostics [152] [153] [154] [155] [156]. For example, Seshadrinath, et al. [153] successfully use a K-NN to classify vibration features extracted using the wavelet transform for the diagnosis of induction machine interturn faults with high accuracy. D. H. Pandya, et al. [154] find a weighted K-NN classifier to outperform other classification techniques in classifying features (including RMS, crest factor, kurtosis and others) for diagnosing bearing faults. J. Tian, et al. [155] use K-NN for diagnosis of bearing faults from kurtosis based vibration features.

The advantage of K-NN as a classifier is simplicity of the technique, as no complex training process is required, and there is only on parameter to alter. In addition, the algorithm is easily adaptable when new training data becomes available, and is easily implementable in incremental learning situations [140]. However, classifying data can be slow, especially when large training datasets are used [140]. This means that the algorithm is less suitable than SVMs and decision trees for use in online systems.

4.3 Deep Learning

While feature-based diagnostic approaches have been proven to be an effective method of fault diagnosis in a number of different applications [29] they have a number of disadvantages.

Firstly, developing a feature-based diagnostics model for a large and complex system, with many different interconnected components, can be a laborious challenge. Rotating components often have a number of different fault modes, each of which with unique fault characteristics (discussed in section 4.1, p. 83) that need to be understood in order to select the most appropriate feature-extraction method (of which there are a number to choose from).

In addition, detailed specifications of the machinery are commonly required to calculate known fault frequencies required in many feature-extraction methods. This may be problematic to source for all components in a large system, where components have been sourced from multiple different manufacturers. Also, some fault frequencies, such as natural resonances, may not be possible to determine without detailed analysis of vibration data after a fault has occurred.

Deep learning provides an alternative solution to feature-based methods. The aim of deep learning approaches is to replace feature-extraction processes with large, multi-layered neural network architectures that automatically learn features from data, optimised to best separate different classes of data [157]. This can provide an advantage to rotating machine diagnostics where labelled training data is available, replacing the intensive manual analysis of data required to develop feature-based methods.

Deep learning has recently been applied in various condition monitoring applications, showing positive results. Sparse coding techniques applied to raw vibration data for feature learning have been used for bearing fault detection by Liu, et al. [158] and Martin del Campo & Sandin [159]. Catterson & Sheng [160] used deep neural networks to diagnose partial discharge data in high voltage assets. Qiao & Xun [161] used deep autoencoder networks with support vector regression to estimate the health state in fuel cell systems.

Deep autoencoder networks have also been proven to be an effective method for classifying rotating machine faults. Verma, et al. [162] trained a deep network to diagnose faults in air compressors from acoustic data. Junbo, et al, [163] used a deep autoencoder network to diagnose bearing faults using coefficients obtained through the wavelet transform, obtaining classification accuracies between 91.25% and 99.62% in multiple loading scenarios. Jia, et al. [164] trained a deep autoencoder network with five feature layers to diagnose bearing and planetary gearbox faults using spectral data obtained through the Fourier Transform. This approach provided high classification accuracy of 99.61% and 100% for bearing faults.

4.3.1 Artificial Neural Networks

The foundations of deep learning come from Artificial Neural Networks (ANNs). ANNs are structures that mimic the learning functionality of the brain, connecting multiple individual processing units (known as neurons or perceptrons) together allowing the network to learn complex nonlinear functions [128].

Deep learning can be performed through a number of different configurations of ANN, including stacked autoencoder networks, convolutional neural networks, deep belief networks and recurrent neural networks [89]. This work has focussed on applying stacked autoencoder networks, based upon feed forward neural networks, to vibration data for rotating machine diagnostics.

4.3.1.1 Feed Forward Neural Networks

A feed forward neural network (FFNN) is an artificial neural network where data flows in one direction, connecting input data parameters to output data parameters through a number of hidden neurons (also known as perceptrons) [128]. By tuning the weights $\boldsymbol{W}$ and biases $\boldsymbol{b}$ associated with each of the neurons within a network, these structures are capable of learning complex nonlinear functions for both regression and classification.

Each neuron in the network performs two functions on data, Figure 4.35. Firstly, all input parameters x_i are multiplied by a weight W_i and summed together with a bias parameter b (4.56). Secondly, a function is then applied as in (4.57) to calculate the activation a of each neuron.

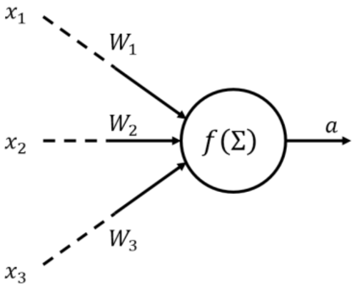


Figure 4.35 - Neuron and synapse

$$z = \sum_{i=1}^{n} W_i x_i + b \tag{4.56}$$

$$a = f(z) = f\left(\sum_{i=1}^{n} W_i x_i + b\right) \tag{4.57}$$

This activation function f is commonly chosen as a sigmoid (4.58), tanh (4.59) or rectified linear function (4.60) for deep learning applications [165].

$$f(z) = \frac{1}{1+e^{-z}} \tag{4.58}$$

$$f(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{e^{2z}-1}{e^{2z}+1} \tag{4.59}$$

$$f(z) = \max(0, z) \tag{4.60}$$

Figure 4.36 shows the structure of a simple FFNN, where input data variables $\boldsymbol{x}$ are connected to output data $\boldsymbol{y}$ through a layer of hidden neurons. The calculation of the activation of each neuron (as in (4.56) and (4.57)) can be performed for each layer of the network using matrix calculations in a process known as the feed forward pass [166], as in (4.61) to (4.64). The activations of the neurons in the output layer of the network are used to estimate output parameters $\boldsymbol{y}$ (4.64).

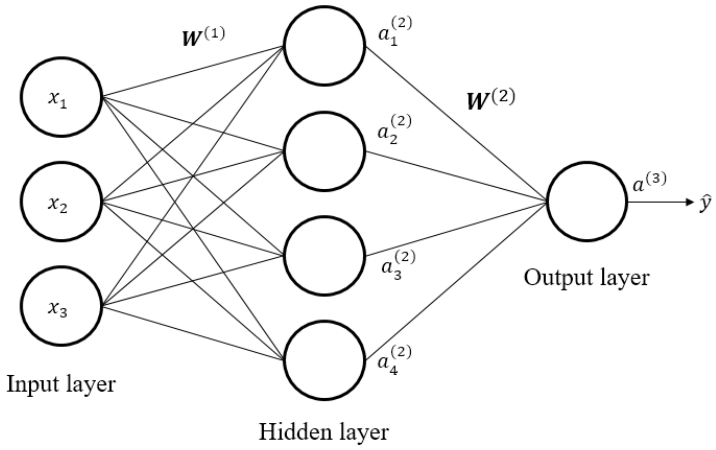


Figure 4.36 - An example of a simple Feed Forward Neural Network

$$\boldsymbol{z}^{(\boldsymbol{2})} = \boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)} \tag{4.61}$$

$$\boldsymbol{a}^{(2)} = f\left(\boldsymbol{z}^{(2)}\right) = f\left(\boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)}\right) \tag{4.62}$$

$$\boldsymbol{z}^{(3)} = \boldsymbol{W}^{(2)}\boldsymbol{a}^{(2)} + \boldsymbol{b}^{(2)} \tag{4.63}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{a}^{(3)} = f\left(\boldsymbol{z}^{(3)}\right) = f\left(\boldsymbol{W}^{(2)}\boldsymbol{a}^{(2)} + \boldsymbol{b}^{(2)}\right) \tag{4.64}$$

4.3.1.2 Training a Neural Network

To learn a complex nonlinear function, neural network weight $\boldsymbol{W}$ and bias $\boldsymbol{b}$ parameters must be optimised through a training process to minimise the error between true outputs $\boldsymbol{y}$ and the estimated outputs $\hat{\boldsymbol{y}}$. This error can be expressed through a cost function, e.g. a quadratic function [128] as in (4.65), where K is the number of training examples.

$$J = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}(y_k - \hat{y}_k)^2 \tag{4.65}$$

The network's error can be minimised by finding parameters which cause the cost function to decrease, i.e. parameters $\boldsymbol{W}$ and $\boldsymbol{b}$ that have a negative gradient with respect to cost function J. This gradient can be represented through the partial derivatives $\frac{\partial J}{\partial \boldsymbol{W}}$ and $\frac{\partial J}{\partial \boldsymbol{b}}$, and are found using the back propagation algorithm [167].

The back propagation algorithm involves calculating the error between true outputs $\boldsymbol{y}$ and estimated outputs $\hat{\boldsymbol{y}}$ and propagating this backwards through each layer of the network to find the effect of this error on each neuron's activation a. The back propagation algorithm can be summarised as below [166]:

1. Perform feed forward pass (as in (4.61) to (4.64)) to find all neuron activations $\boldsymbol{a}$ in all layers.
2. Calculate error term δ at output layer (layer L) from true output values $\boldsymbol{y}$ (4.66).

$$\delta^{(L)} = -(y - \hat{y}) \circ f'\left(z^{(L)}\right) \tag{4.66}$$

3. Working backwards, calculate δ in all remaining layers in the network (4.67).

$$\delta^{(l)} = \left(\left(\boldsymbol{W}^{(l)}\right)^T \delta^{(l+1)}\right) \circ f'\left(z^{(l)}\right) \tag{4.67}$$

4. Compute partial derivatives from error terms (4.68) and (4.69).

$$\frac{\partial}{\partial \boldsymbol{W}^{(l)}} J(\boldsymbol{W}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{y}) = \delta^{(l+1)}\left(a^{(l)}\right)^T \tag{4.68}$$

$$\frac{\partial}{\partial \boldsymbol{b}^{(l)}} J(\boldsymbol{W}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{y}) = \delta^{(l+1)} \tag{4.69}$$

Once partial derivatives are calculated a number of different training algorithms can be employed to minimise the prediction error of the network. One of the simplest techniques is gradient descent, where $\boldsymbol{W}$ and $\boldsymbol{b}$ are tuned iteratively as in (4.70) and (4.71), with η as some scalar constant, until convergence is reached [168].

$$\boldsymbol{W}(t) = \boldsymbol{W}(t-1) - \eta\frac{\partial J}{\partial \boldsymbol{W}} \tag{4.70}$$

$$\boldsymbol{b}(t) = \boldsymbol{b}(t-1) - \eta\frac{\partial J}{\partial \boldsymbol{b}} \tag{4.71}$$

In more advanced training algorithms, η becomes an estimate of the inverse Hessian matrix containing second derivatives of the cost function [168], e.g. scaled conjugate gradient [169] or Quasi-Newton methods such as the Marquardt-Levenberg algorithm [80].

4.3.2 Deep Learning through Stacked Autoencoders

In this work, deep learning is performed through stacked autoencoder networks. These networks automatically learn features from input data through multiple interconnected FFNN layers known as autoencoders [74]. A final connected layer on the network then performs classification.

These networks are trained in two stages [170]. The first stage tunes the parameters of each autoencoder layer sequentially to learn features from the structure of input data. The second stage then optimises the features learned by the network to give the best separation between output classes and maximise classification performance.

4.3.2.1 Stacked Autoencoder Networks

Autoencoders are a form of neural network used for unsupervised learning [74], where networks are trained with only input data $\boldsymbol{x}$ and no output data $\boldsymbol{y}$. Their function is to replicate input data $\boldsymbol{x}$ at their output through a layer of hidden neurons, Figure 4.37. The network therefore learns a function $h_{\boldsymbol{W},\boldsymbol{b}}$ to best approximate the input, (4.72).

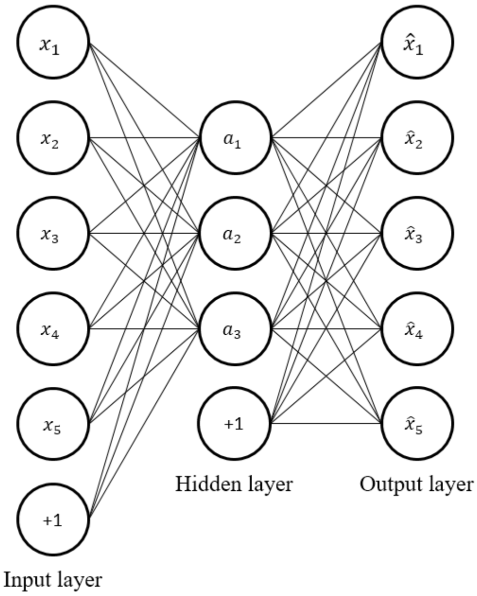


Figure 4.37 - Autoencoder neural network

$$\hat{\boldsymbol{x}} = h_{\boldsymbol{W},\boldsymbol{b}}(\boldsymbol{x}) \approx \boldsymbol{x} \tag{4.72}$$

By limiting the number of hidden neurons, the network is forced to learn a compressed representation of the input data. Hidden neurons can therefore be seen as representing features, learned from structure within the input data [74].

Deep neural networks can be constructed by stacking many autoencoder layers on top of each other [171]. This is shown in Figure 4.38, where L autoencoder layers are connected together before a final classification layer. Having multiple hidden layers allows the network to further decompose the structure of input data, with the complexity of features increasing in later layers of the network [172].

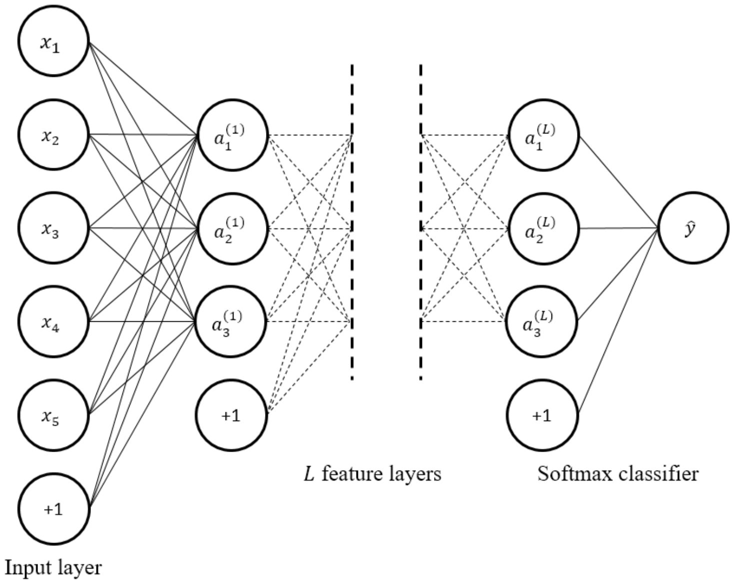


Figure 4.38 - Stacked autoencoder classification neural network

4.3.2.2 Training a Deep Network

Deep autoencoder networks can be trained in two stages: unsupervised pre-training and supervised fine-tuning.

Pre-training involves training each autoencoder layer by layer in a process known as greedy layer-wise training [170] [173]. The first autoencoder layer is trained on raw input data $\boldsymbol{x}$, learning feature representations in the hidden layer represented by $\boldsymbol{W}^{(1)}$ and $\boldsymbol{b}^{(1)}$. The activations of each hidden neuron in the first layer $\boldsymbol{a}^{(1)}$ are then used as the inputs to the second autoencoder layer. This layer is then trained to reconstruct its input, learning more complex features represented through $\boldsymbol{W}^{(2)}$ and $\boldsymbol{b}^{(2)}$. This process is repeated for L autoencoder layers.

Each training stage can use any training algorithm, however the cost function in (7.65) is altered as in (4.73) for all autoencoder layers, introducing a weight decay term Ω_{weights} and a sparsity term Ω_{sparsity} [174], where K is the number of training examples.

$$J = \left[\frac{1}{K} \sum_{j=1}^{K} \frac{1}{2} \left\| \boldsymbol{y}_j - h_{\boldsymbol{W},\boldsymbol{b}}(\boldsymbol{x}_j) \right\|^2 \right] + \Omega_{\text{weights}} + \Omega_{\text{sparisty}} \tag{4.73}$$

The weight decay term $\Omega_{weights}$ limits the magnitude of weights to prevent overfitting [174] and is calculated as in (4.74), where L is the number of autoencoder layers, $N^{(l)}$ is the number of neurons in layer l, K is the number of training examples and λ is a weight decay parameter controlling the influence of this term on the cost function.

$$\Omega_{weights} = \frac{\lambda}{2} \sum_{l}^{L} \sum_{i}^{N^{(l)}} \sum_{j}^{K} \left(\boldsymbol{W}_{ij}^{(l)} \right)^2 \tag{4.74}$$

The sparsity term $\Omega_{sparsity}$ encourages sparse activations of hidden neurons within the network, allowing more diverse features to be discovered even where autoencoders have a large number of hidden units. It does this by applying the constraint $\hat{\rho}_i = \rho$, where $\hat{\rho}_i$ is the average activation of hidden neuron i and ρ is a sparsity parameter set typically close to zero (e.g. 0.05). The average activation $\hat{\rho}$ of hidden neuron i in a network can be expressed as in (4.75).

$$\hat{\rho}_i = \frac{1}{K} \sum_{j=1}^{K} f\left(\boldsymbol{W}_{ij}^{(l)T} \boldsymbol{x}_j + \boldsymbol{b}_i^{(l)} \right) \tag{4.75}$$

This constraint is enforced through Kullback-Leibler divergence [175], shown in (4.76), where β is a parameter controlling the influence of the sparsity term in the cost function.

$$\Omega_{sparsity} = \beta \sum_{i=1}^{N^{(l)}} \mathrm{KL}(\rho || \hat{\rho}_i) = \beta \sum_{i=1}^{N^{(l)}} \rho \log\left(\frac{\rho}{\hat{\rho}_i} \right) + (1 - \rho) \log\left(\frac{1 - \rho}{1 - \hat{\rho}_i} \right) \tag{4.76}$$

Once each autoencoder layer is trained, the final set of neuron activations $\boldsymbol{a}^{(L)}$ are used as inputs for training a softmax classifier layer with target outputs $\boldsymbol{y}$. This type of neural network classifier is trained using a cross-entropy cost function [173] as in (4.77).

$$J = \frac{1}{K} \sum_{j=1}^{K} \boldsymbol{y}_j \log\left(h_{\boldsymbol{W},\boldsymbol{b}}(\boldsymbol{x}_j) \right) + \left(1 - \boldsymbol{y}_j \right) \log\left(1 - h_{\boldsymbol{W},\boldsymbol{b}}(\boldsymbol{x}_j) \right) \tag{4.77}$$

Once all weights $\boldsymbol{W}$ and bias $\boldsymbol{b}$ values are initialised in pre-training, the network is the fine-tuned [176]. This process involves training the fully connected network as a whole using supervised learning, where both input data $\boldsymbol{x}$ and output data $\boldsymbol{y}$ are used, altering $\boldsymbol{W}$ and $\boldsymbol{b}$ to improve the classification accuracy.

4.4 Case Studies

Two case studies have been examined in this work, demonstrating both feature-based diagnostics and deep learning diagnostics applied in practice for the detection of rotating machine faults.

The first case study compares a feature-based diagnostic method to a deep learning approach for the diagnosis of a generator fault identified within accelerometer data recorded from the HS1000. The feature-based method involved several stages, including a Vold-Kalman filter for order tracking, RMS for building a feature set, normalisation through the Z-score standardisation and finally a classification technique. Multiple classification methods were compared, including variations of SVMs, decision trees and K-NN. Deep learning was performed through a stacked autoencoder neural network.

The second case study details the use of a stacked autoencoder neural network to detect multiple different faults from a gearbox test rig. This dataset was made available as part of the PHM 2009 data challenge [177] and included vibration data from faults in gear stages, bearings and shafts, recorded over multiple loading conditions.

4.4.1 Methodology

4.4.1.1 Feature-based Diagnostics

Feature-based diagnostics was performed as in Figure 4.39.

The Vold-Kalman filter (section 4.2.1.1.2, p. 103) was first used to extract characteristic fault frequencies at varying rotation speeds. A feature set was then generated by taking the RMS (section 4.2.2.1, p. 106) of each frequency component over an N second window.

Before classification, features were scaled using Z-score standardisation [72], detailed in (4.78), where $\bar{x}$ and σ are the mean and standard deviation of feature x respectively. This transformation gives the data zero mean with a standard deviation of one.

$$Z' = \frac{x - \bar{x}}{\sigma} \tag{4.78}$$

Scaled features are then classified, comparing SVM, decision tree and k-nearest neighbour techniques (described in section 4.2.3, p. 113). The performance of each technique was measured through the classification accuracy as a percentage as calculated in (4.79), where N_{samples} is the total number of test samples and N_{corect} is the number of correctly classified test samples.

$$\text{accuracy (\%)} = \left(\frac{N_{\text{correct}}}{N_{\text{samples}}}\right) \times 100 \tag{4.79}$$

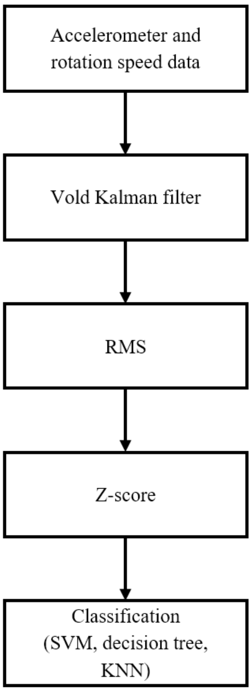


Figure 4.39 - Feature-based diagnostics process

4.4.1.2 Deep Learning

Stacked autoencoder networks (described in section 4.3.2, p. 122) were trained as fault classifiers using narrow spectrograms generated from raw vibration data as inputs. This gives a time-frequency representation of vibration data, allowing the network to learn from both stationary and non-stationary fault signals.

These deep networks first learned features from vibration data using unsupervised learning through a number of autoencoder neural network layers. These layers were then connected to a softmax classification layer, trained using supervised learning to identify faults. The fully connected network was then retrained using the fine tuning process (section 4.3.2.2, p. 123) to further tune the learned features to improve classification accuracy.

Input narrow spectrogram data, referred to as spectrogram ‘slices’, were generated using the Short Time Fourier Transform as in section 4.2.2.3.1, p. 108. This data represented the change in frequency response over short periods of time. The width of these slices was set to N seconds, matching the size of the RMS window in the feature-based diagnostics mythology. Features learned by the deep networks would be representative of individual frequency components relating to component faults obtained from the spectrogram data.

Network training data was randomly sampled from 70% of the spectrogram data. The remaining 30% of the data was used for testing. Networks of different configuration (altering the number of autoencoder layers and hidden neurons) were evaluated using the percentage of samples correctly classified, calculated as in (4.79).

Despite the neural network being primarily a ‘black-box’ technique, features learned by autoencoder layers at each hidden neuron can be visualised. The input $\boldsymbol{x}_j$ which causes hidden neuron i to have maximum activation can be used to represent the feature learned by the neuron. This is dependent on the weight matrix $\boldsymbol{W}$ of each neuron, and is calculated as in (4.80) [166].

$$\boldsymbol{x}_j = \frac{\boldsymbol{W}_{ij}^{(1)}}{\sqrt{\sum_{j=1}^{K}\left(\boldsymbol{W}_{ij}^{(1)}\right)^2}} \tag{4.80}$$

4.4.2 Case Study 1: HS1000 Generator

In this case study, the feature-based methodology was compared to the deep learning methodology for the detection of a suspected induction generator fault observed within vibration data recorded from the HS1000 in operation. This data was captured during two periods of operation, where the turbine experienced turbulent loading and variable rotation speeds.

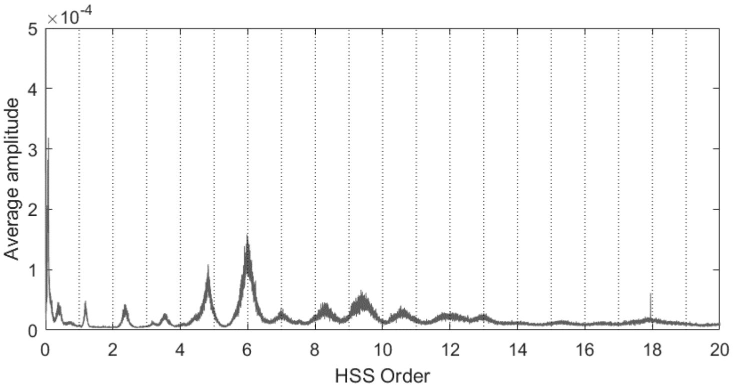
4.4.2.1 Fault Case

High resolution accelerometer data was available for the HS1000 tidal turbine (as discussed in chapter 2), recorded through a tri-axial accelerometer sensor fixed to the nacelle of the turbine, adjacent to the generator, with a sample rate of 2 kHz. A number of vibration measurements were taken over short windows (approximately 5 minutes) during two separate deployments of the turbine with a total period of over 10 months.

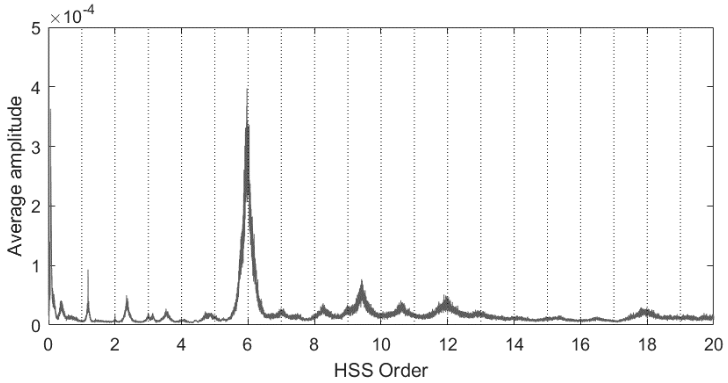
Measurements were taken during a range of different operating and weather conditions, where the turbine experienced constantly variable loading and rotation speed due to changing tidal flow rates and turbulence.

During the turbine’s second deployment, the turbine displayed a change in response causing increased vibrations within the turbine’s generator. This can be observed in Figure 4.40, showing the average spectrum of vibration data for each deployment period. Computed resampling [101] has been used to normalize the frequency spectrum to represent rotational orders of the high speed shaft (HSS).

A clear increase in vibration can be observed at the 6th order of the HSS, which may relate to the generator’s poles. High vibrations were also identified at 100 Hz (a harmonic of the line frequency) and at harmonics of the generator rotation frequency during different loading conditions. These frequencies were most consistent with an eccentric rotor fault (section 4.1.3.8, p. 98). However, strip-down information was not available to confirm this observation following decommissioning of the turbine.



(a) HS1000 operation between – baseline healthy response



(b) HS1000 operation between – generator fault

Figure 4.40 - Averaged computer resampled order spectrum (high speed shaft) for two separate deployments of the HS1000 tidal turbine detailing a generator fault

4.4.2.2 Results

Feature-based classification was performed as in 4.4.1.1., where features representing the following fault frequencies characteristic of the observed generator fault were extracted under varying rotation speeds using the Vold-Kalman filter:

- Generator shaft rotation frequency and 2nd harmonic
- Generator poles vibration frequency and 2nd harmonic
- Line frequency 2nd harmonic (100 Hz)

The RMS of each component was calculated over a 1 second window, before applying the Z-score transformation.

Table 4.1 details the results of the feature-based diagnosis method, where the classification accuracies of multiple different classification methods (including SVMs with different kernel functions, decision tree and K-NN) were calculated on test data as in (4.79). This feature-based method gives positive results, with high classification accuracies of above 93% for all classification methods. A cubic SVM classifier gives the best classification accuracy of 96.86%.

Deep learning was performed using the methodology described in 4.4.1.2. Spectrogram data was generated by performing 10 overlapping STFTs every second, with the frequency scale truncated to 500 Hz, removing redundant frequency information and reducing the dimensionality of input data for more computationally efficient training.

Table 4.2 details the classification accuracies of the deep learning diagnosis method for networks with varying numbers of autoencoder layers and hidden neurons after both pre-training and fine-tuning stages. These results show the effectiveness of deep learning, where classification accuracy exceeds the feature-based method, reaching 100% in many network configurations.

For each network configuration, increasing the number of hidden units increases the accuracy in each training stage. However, increasing the number of autoencoder layers, allowing deep networks to further deconstruct input data into more complex features, actually decreases the classification accuracy in the pre-training stage. This indicates that the structure of the input data is of limited complexity and can be modelled more effectively with smaller networks.

Table 4.1 - Feature-based classification accuracy for identification of a generator fault on the HS1000

Classification Method		**Testing accuracy**
SVM	Linear SVM	93.01 %
	Quadratic SVM	96.06 %
	Cubic SVM	96.86 %
Decision Tree	20 maximum nodes	95.96 %
	50 maximum nodes	96.85 %
	100 maximum nodes	96.39 %
K-NN	$k = 1$	96.42 %
	$k = 10$	96.76 %
	$k = 100$	95.59 %

Table 4.2 - Deep learning classification accuracy for identification of a generator fault on the HS1000

Number of autoencoder layers	Number of Hidden Units	Pre-training accuracy	Fine-tuning accuracy
1	5	89.73%	99.97%
	10	96.59%	99.97%
	50	99.56%	100%
	100	99.85%	100%
2	5	68.71%	99.97%
	10	72.32%	99.97%
	50	73.73%	100%
	100	73.86%	100%
3	5	43.81%	99.97%
	10	56.19%	99.97%
	50	73.13%	100%
	100	73.35%	100%
4	5	54.99%	99.97%
	10	54.99%	100%
	50	56.78%	100%
	100	65.23%	100%

4.4.2.3 Feature Visualisation

Figure 4.41 displays $\boldsymbol{x}_j$, calculated as in (4.80), for selected hidden neurons in the first autoencoder layer with the highest activation for faulty conditions. Bright areas in each figure represent higher dependency on input data at that frequency (i.e. the neuron is looking for data at that frequency to make a classification) and darker areas represent negative dependency (i.e. the neuron is looking for the absence of data at that frequency). Areas of mid grey represent low dependency of data at that frequency (i.e. the neuron is not concerned with data at these frequencies).

Analysis of the features learned by each neuron shows a series of bright areas at frequency components relating to vibration due to the generator’s poles at different rotation speeds. This indicates the network has decomposed input data to automatically identify fault signatures of the generator without prior engineering knowledge, using the activation of these features for classification.

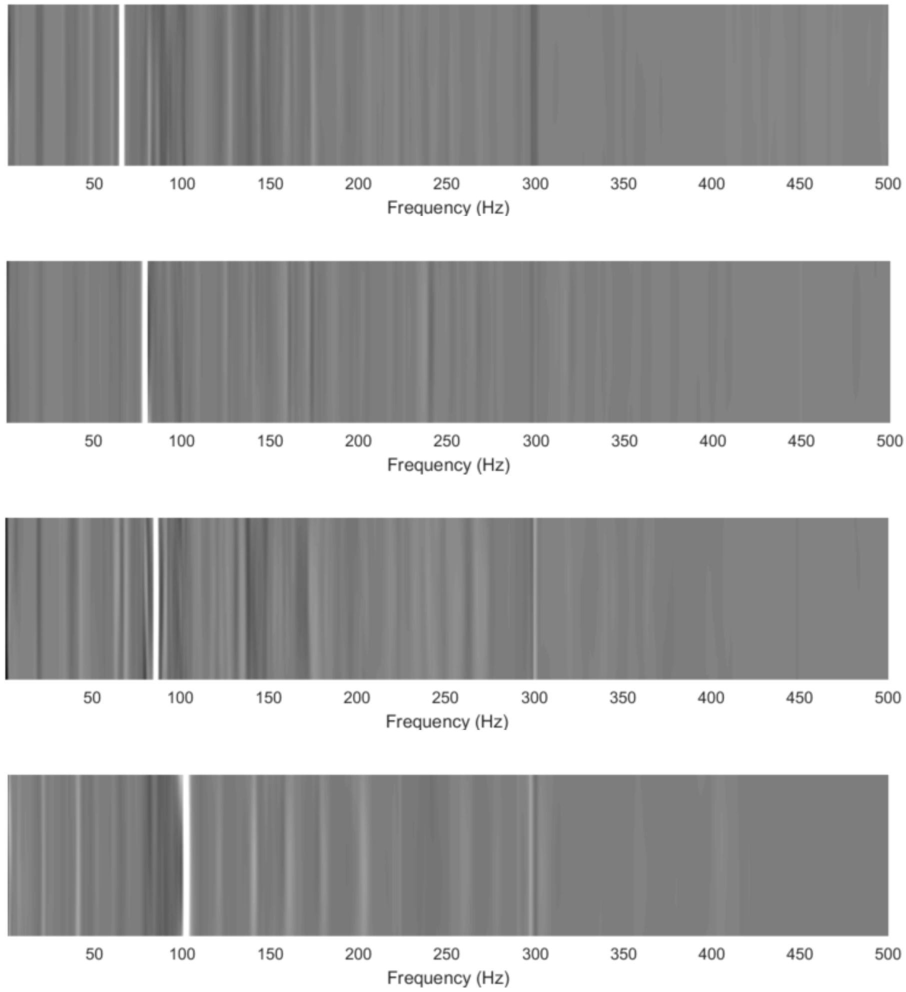


Figure 4.41 - Features learned by stacked autoencoder network showing generator poles vibration deconstructed from vibration data at different rotation speeds

4.4.3 Case Study 2: Gearbox Test Rig

This case study examines the use of stacked autoencoder networks for diagnosing multiple rotating machine failures from the same system under variable loading conditions. Data was sourced from a gearbox test rig, where faults were introduced in a number of different components, including gear stages, bearings and shafts [177].

Such faults can produce both stationary and non-stationary vibration signals. In addition, no single feature-based diagnostic technique best deals with these different types of vibration, as no single process is effective at isolating all types of signal associated with rotating machine fault modes [29]. This case study would therefore test the ability of stacked autoencoder networks to isolate and learn features from a combination of stationary and non-stationary signals simultaneously without the need for multiple different signal processing methods for separate faults.

4.4.3.1 Gearbox Dataset

The data used in this case study was released as part of the PHM Society's 2009 PHM Challenge competition [177]. The aim of the dataset is to demonstrate the vibrations exhibited by a gearbox running under a number of different operational settings and under a number of different interacting fault modes.

This dataset detailed a gearbox test rig, configured as in figure 4.42, consisting of three shafts, each with two supporting bearings, and two gear stages. This gearbox was designed to be reflective of a generic industrial gearbox. Each gear stage was run in two configurations. The first configuration used spur gears with 32 gear teeth and 96 pinion teeth in the first stage, and 48 gear teeth and 80 gear teeth in the second stage. The second configuration used helical gears with 16 gear teeth and 48 pinion teeth in the first stage, and 24 gear teeth and 40 pinion teeth in the second gear stage. The first gear stage reduced the speed of an input shaft to an idler shaft by a factor of 3. The second stage reduced the speed of the idler shaft by a factor of 1.667, giving a total rotation speed reduction ratio of 5.

For each configuration, the gearbox run under several different fault modes, listed in Table 4.3. Fault modes were introduced manually, replacing healthy gear and bearing components with faulty components manufactured within the laboratory. The test rig was also run under a number of different operational settings for 4 second intervals. This included varying the speed between 30, 35, 40, 45 and 50 Hz, under both high a low loading. The run time and load were chosen in order to not introduce any additional fault progression under each operating condition.

Vibration data under each condition was recorded by two accelerometers located at the input and output sides of the gearbox respectively, sampling data at 66.667 kHz for 4 second intervals. A tachometer was also used to measure the rotation speed of the input shaft during these time intervals.

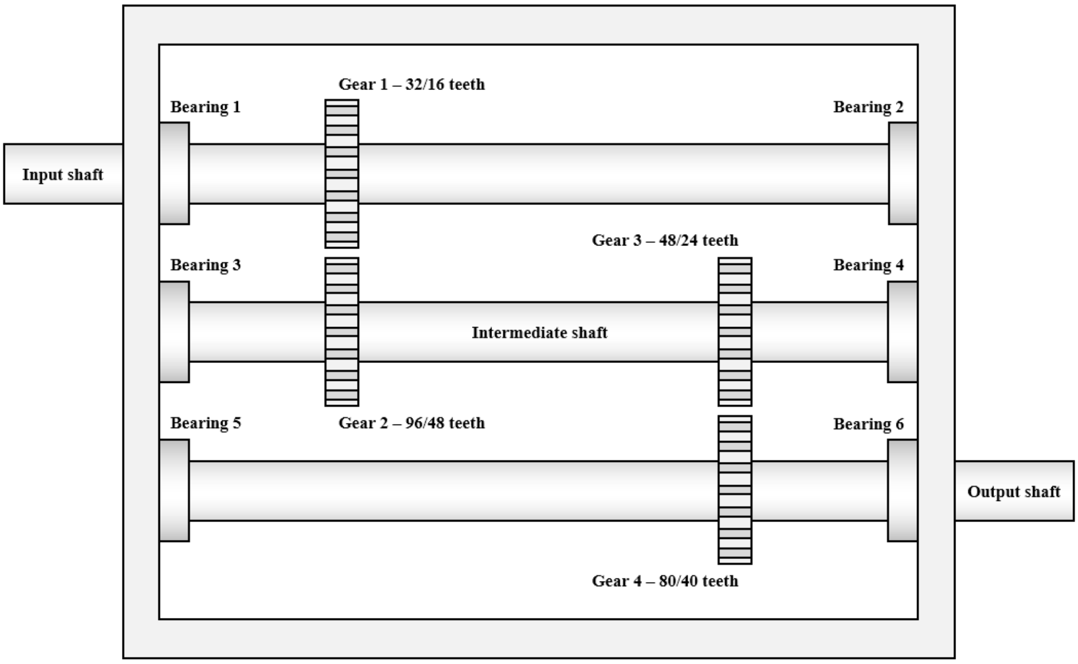


Figure 4.42 - Gearbox test rig apparatus, adapted from [177]

Table 4.3 - Independent fault modes in gearbox test rig dataset

Gear Type	Component	Fault
Spur gear	Gear	Chipped gear tooth
		Eccentric gear
		Broken gear tooth
	Bearing	Ball fault
		Inner raceway fault
	Shaft	Imbalance
		Keyway sheared
Helical gear	Gear	Chipped gear tooth
	Bearing	Combination
	Shaft	Bent shaft

4.4.3.2 Results

Spectrogram data was generated by performing 10 overlapping STFTs every 0.01 seconds. This window length was reduced from the previous case study to best capture non-stationary effects in the vibration signal occurring at the maximum rotation speed of 50 Hz. The frequency scale of each spectrogram slice was truncated to 10 kHz to further reduce the dimensionality of data.

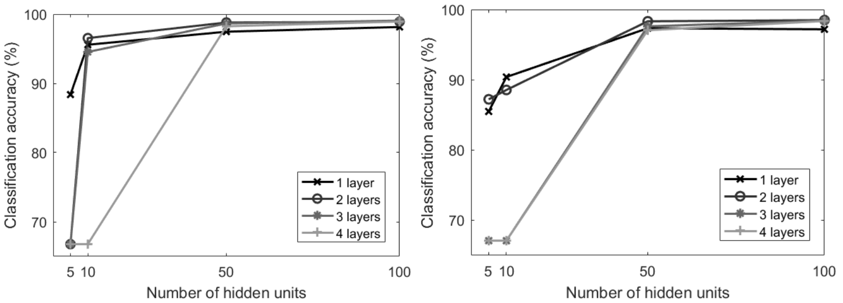
A separate deep autoencoder network was constructed to detect each fault mode in Table 4.3, each trained as a binary classifier. The number of autoencoder layers and the number of hidden units in each layer were varied to see the effect on classification performance.

Table 4.4 lists the maximum classification accuracy recorded for each fault. Figure 4.43 shows the classification accuracies for each fault under different network configurations. Results showed high classification accuracies for each fault type, ranging from 98.6935% accuracy for detection of a bearing inner raceway fault to 100% accuracy for the detection of shaft keyway shearing. As in the previous case study, the classification accuracy was observed to increase as the number of hidden units in each autoencoder layer was increased.

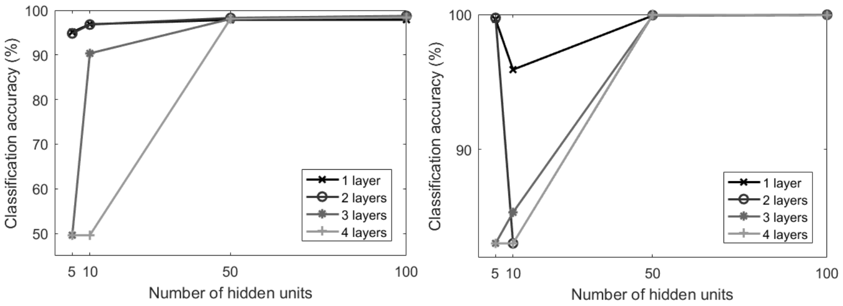
Figure 4.43 shows classification accuracy was also improved by small amounts for most fault cases when the number of hidden layers was increased. This implies that there is some additional complexity within the input data, potentially due to the combination of both stationary and non-stationary vibration signals from gear stage and bearing faults.

Table 4.4 - Maximum classification accuracies of deep learning for each independent fault mode

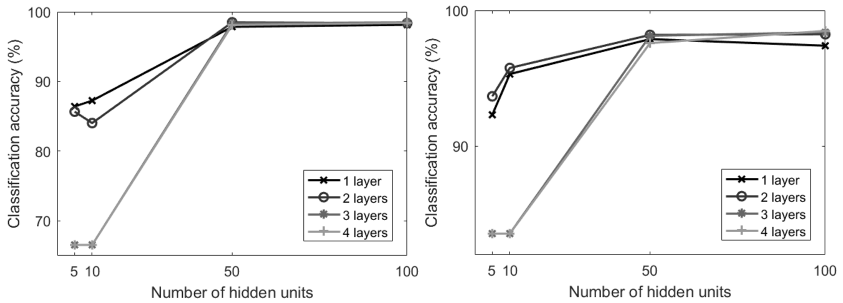
Gear type	Fault	Maximum classification accuracy
Spur gear	Chipped gear tooth	99.06%
	Eccentric gear	98.80%
	Broken gear tooth	98.80%
	Ball fault	99.97%
	Inner raceway fault	98.69%
	Imbalance	98.85%
	Keyway sheared	100.00%
Helical gear	Chipped gear tooth	99.18%
	Combination	99.51%
	Bent shaft	99.09%



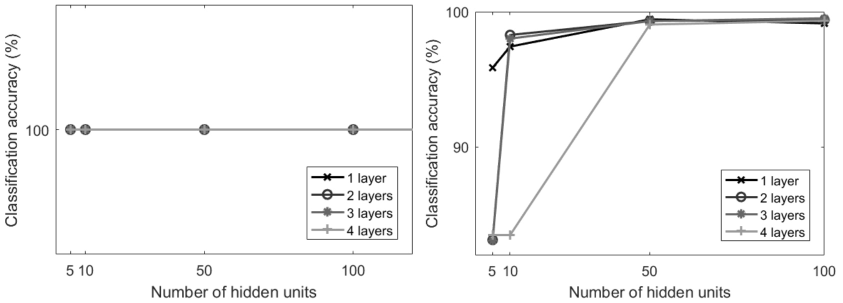
(a) Chipped gear tooth (spur)
(b) Eccentric gear (spur)



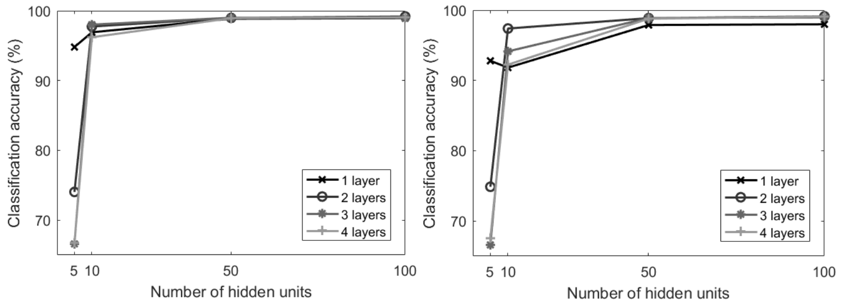
(c) Broken gear tooth (spur)
(d) Bearing ball fault (spur)



(e) Bearing inner raceway fault (spur)
(f) Shaft imbalance (spur)



(g) Shaft keyway sheared (spur)
(h) Chipped gear tooth (helical)



(i) Combination bearing fault (helical)
(j) Bent shaft (helical)

Figure 4.43 - Classification accuracies of each network configuration for each fault mode

4.4.3.3 Feature Visualisation

Features learned by each network were visualised as in (4.79), where bright areas represent higher dependency on input data at that frequency, dark areas represent high negative dependency and mid grey areas represent low dependency at that frequency for classification.

Examples of features with high activations during faulty conditions for a number of different networks are shown in Figure 4.44(a) to (f), where bright areas in each figure show that networks appear to have learned representations of both known fault frequencies and natural resonances.

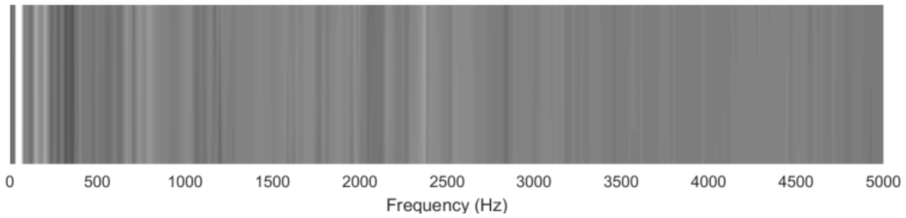
Figure 4.44(a) shows a feature representing the rotation speed of the input shaft at 50 Hz, where rotation frequency is a common feature of rotating component fault modes.

Figure 4.44(b) shows a feature with high dependency for BPFI frequency, a frequency linked to a fault within a bearing's inner raceway. Dark areas in the figure surrounding this frequency represent high negative dependency, where the network may be filtering out adjacent frequency components or noise present in the input data but not related to this particular fault mode.

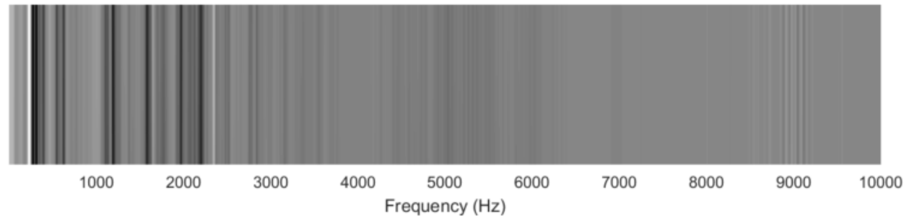
Figure 4.44(c) and (d) show features with high dependency for a gear mesh frequency and a gear mesh frequency with multiple harmonics, which are frequencies related to gear stage faults.

Figure 4.44(e) and (f) show features with high dependency at high frequencies (above 5 kHz). These frequencies are believed to be from non-stationary natural resonances within components of the gearbox, invoked by gear and bearing faults respectively.

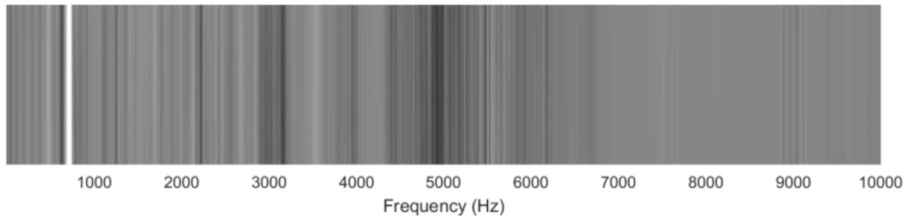
These results show that deep networks trained with vibration spectrogram data can automatically learn vibration fault signatures purely from structure within the input data and with no prior engineering knowledge of the system or its response to failure. Each network also learns a combination of many other frequencies, not necessarily linked to known fault frequencies derived from engineering principles, but potentially representative of measured fault signals within the system.



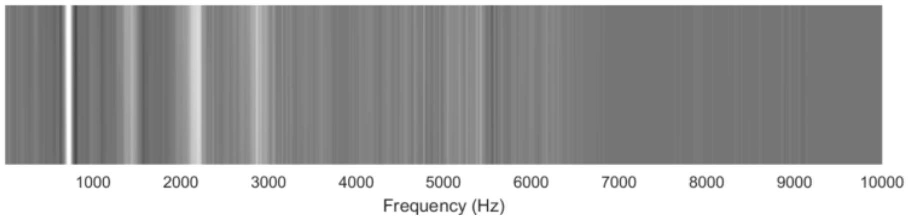
(a) Input shaft rotation speed



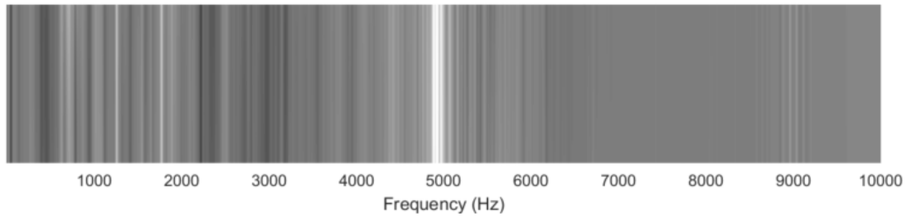
(b) Bearing BPFI



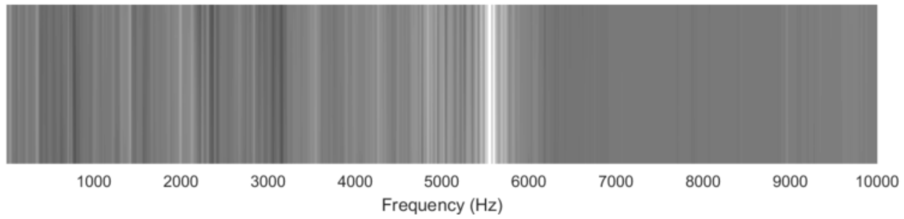
(c) Gear mesh



(d) Gear mesh and multiple harmonics



(e) High frequency natural resonance



(f) High frequency natural resonance

Figure 4.44 - Examples of features learned by stacked autoencoder networks for different fault modes

4.5 Diagnostics in an Intelligent Condition Monitoring System

In chapter 3, anomaly detection was discussed, where CRISP-DM was used to examine available monitoring data to model the normal response of machinery. This is an appropriate first stage in condition monitoring in new applications, where fault data is not readily available, allowing faults to be detected an early stage and allowing operators to act before faults propagate further throughout the machine. However, these methods do not give an implicit classification of the faults that have occurred within machinery.

Figure 4.45 shows how this CRISP-DM approach for anomaly detection can be extended to introduce diagnostic models as fault data becomes available.

The extended model includes feedback after the deployment stage, returning to the data understanding phase once a fault has been detected by an anomaly detection model. This allows diagnostic models to be introduced to the system as faults are detected. The data understanding phase should then be used to identify differences between healthy and faulty data.

A decision stage has been introduced after the data understanding phase. If no fault data is available, anomaly detection models should be constructed, using CRISP-DM as discussed in chapter 3. However, if fault data is available, diagnostic models should also be constructed.

A data preparation stage is introduced after this decision stage for fault data. If feature-based diagnostic methods are being used, this stage will involve signal processing techniques for feature-extraction. If deep learning methods are being used, this stage will involve converting raw sensor data into input data for deep neural networks (e.g. performing the STFT to generate spectrograms). This stage will also include building training and testing datasets, and separating healthy data from fault data.

The diagnostic modelling stage involves training diagnostic models, either classification models for feature-based diagnostics (SVMs, decision trees, etc.) or deep neural networks. Diagnostic modelling is followed by an evaluation stage, where the classification accuracy of models should be determined. This includes feedback, allowing separate models to be compared or updated to improve evaluation results.

Finally, once the performance of diagnostic models is satisfactory, they can be integrated into the intelligent condition monitoring system within the deployment stage.

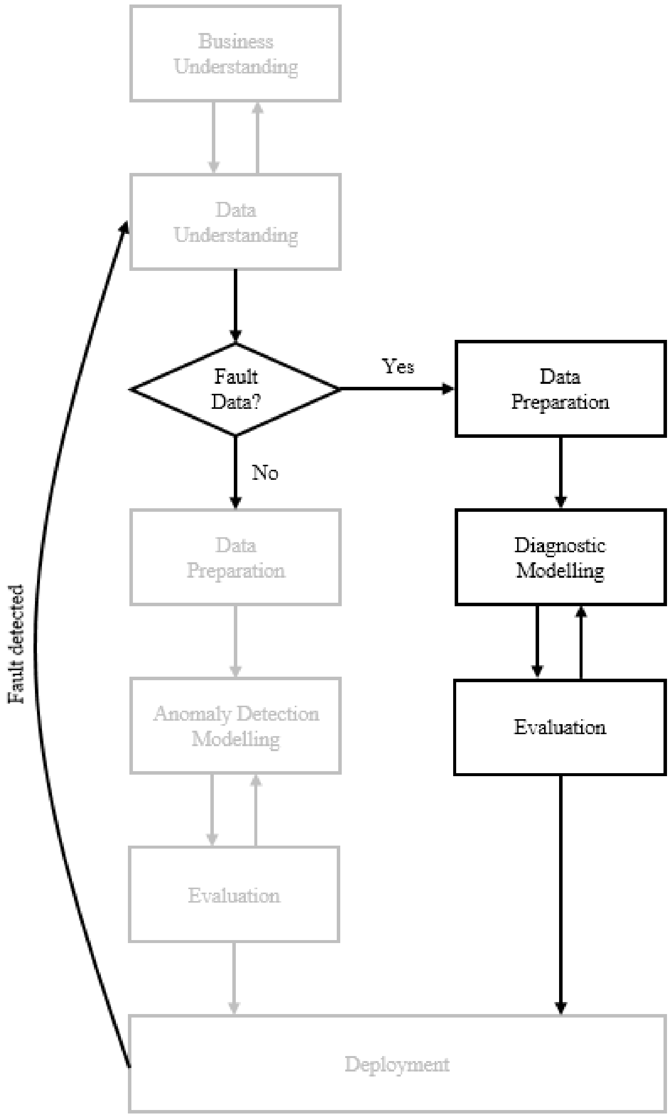


Figure 4.45 - CRISP-DM anomaly detection approach extended for diagnostic modelling

4.6 Summary

This chapter details diagnostic modelling for rotating machines that can be used by an intelligent condition monitoring system to provide operators with automatic monitoring decisions.

First, section 4.1 discusses common fault modes of rotating components (gearboxes, bearings and generators), where high resolution accelerometer data was found to be the most effective parameter in diagnosing rotating machine faults.

Secondly, section 4.2 discusses feature-based methods of automatic fault diagnosis from vibration data, where features are extracted from vibration data representative of fault signals and used as the input to classifiers, trained to make automatic diagnoses. These methods are effective in combining engineering knowledge of the response of rotating machine faults with signal processing and machine learning techniques. However, there is no one technique or process that is most effective at diagnosing all types of fault and developing feature-based diagnostic systems for large machines with many interconnected components, each with a number of unique fault modes, can therefore be labour intensive.

Deep learning is a recently advancing field of machine learning and provides an alternative to feature-based approaches where features are learning automatically through large neural networks. This is discussed in section 4.3, where the stacked autoencoder deep neural networks was identified as a suitable deep learning approach for rotating machine diagnostics.

In section 4.4., two case studies examined both feature-based and deep learning diagnostic methods for detecting faults from accelerometer data. The first case study compared both methods in diagnosing a suspected generator fault within the HS1000. Feature-based diagnostics was performed using the Vold-Kalman filter to isolate characteristic frequencies at varying rotation speed. SVM, decision tree and K-NN classification methods were then compared. Each classification method gave comparable performance, with a cubic SVM providing the highest classification accuracy of 96.86%. Deep learning was then performed through stacked autoencoder networks, trained with spectrogram data obtained through the STFT. This approach was found to out-perform the feature-based method, reaching 100% classification accuracy even for networks with smaller numbers of hidden layers.

The second case study demonstrated this deep learning approach applied to data from a gearbox test rig, identifying multiple different faults from gear stages, bearings and shafts. Deep networks again performed well in this application, giving a classification accuracy of above 98% for all fault modes. Increasing the number of autoencoder layers and hidden neurons in this case, allowing networks to model additional complexity, potentially due to the presence both stationary and non-stationary fault signals, in the input data.

These results show that both feature-based and deep learning approaches are suitable for diagnostic modelling of rotating machines once fault data becomes available. Feature-based learning can be used if there is a good understanding of fault modes and signal processing techniques in order to extract the best features. Once suitable features have been extracted, SVMs appear to give best classification performance, however models can be difficult to interpret and require the selection of an appropriate kernel function. If interpretability is required then decision trees can be used, however these may underperform in comparison to SVMs and over fit the training data.

As an alternative, deep learning can be employed for diagnostics where there is a limited understanding of a system's response to fault modes. Features are learned directly from the data and have been shown to out-perform manually selected features for classification. Additionally, multiple fault modes can be detected using the same method, provided separate models are trained for each type fault. However, adequate training data is required for feature learning and training can require large amounts of computational power.

Finally, CRISP-DM used for anomaly detection, as in chapter 3, was extended for diagnostic modelling. This allows diagnostic models to be incorporated into intelligent condition monitoring systems designed for new applications once fault data becomes available.

Chapter 5

5 Prognostics

The final stage of condition monitoring, following anomaly detection and diagnostics, is prognostics. The aim of prognostics is to predict the time until failure will occur within a component undergoing degradation [30].

Applying prognostics not only requires knowledge of individual fault modes (as in diagnostics) but also requires knowledge of the progression of faults over time. In state of the art prognostic techniques, this knowledge is often inferred from historical data, or from previous examples of failure where past run-to-failure data is available for the system [178].

Implementing prognostics in new systems can therefore be challenging, as historical run-to-failure data is sparsely available. Suitable prognostics methods to predict the remaining health of components must therefore not solely rely on past examples of failure in the current operating environment, and be flexible in incorporating knowledge of failure from other application areas.

In this chapter, the requirements of different prognostic methods are discussed to find the most suitable techniques for making prognostic estimates in new applications with limited past examples of failure.

Model-based degradation techniques are then examined in more detail, in particular Bayesian filtering methods (such as variations of the Kalman filter, and the particle filter), where prognostic estimates are formed from a probabilistic estimate of the current damage state and extrapolated towards a failure threshold using a degradation model.

An approach for implementing model-based prognostics in new systems is then discussed, taking into account the selection of suitable degradation parameters and models. An example of an implementation of the particle filter is then given, following this methodology to estimate the RUL of the HS1000 gearbox from low resolution vibration data under a simulated failure.

Finally, the integration of prognostic modelling into the CRISP-DM framework for developing intelligent condition monitoring methods for new applications is discussed.

5.1 Prognostic Methods

The aim of prognostic methods is to estimate the time until a component will fail after a fault has been detected. This requires knowledge of the fault itself but also knowledge of the rate at which the system will degrade due to the fault over time. This knowledge of the rate of failure can come from different sources, e.g. from historical trends or observations, run-to-failure sensor data from past examples of failure, or prior engineering or physics knowledge of the system used to form a degradation model [30].

Prognostic methods can be therefore split into three categories: time-to-failure analysis, stressor-based and degradation-based [179] where each method forms prognostic estimates based on different forms of input data representative of component degradation over time, as shown in Figure 5.1.

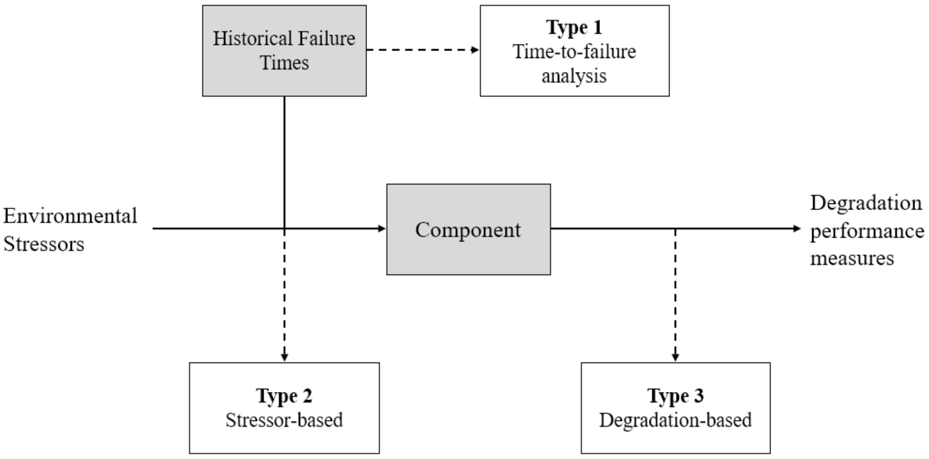


Figure 5.1 - Types of prognostics methods, adapted from [179]

The requirements and assumptions of each of these categories of prognostics is discussed in more detail as follows.

5.1.1 Type 1: Time to Failure Analysis

Time to failure analysis is the most straightforward category of prognostic methods, stemming from traditional reliability analysis. Here, prognostic estimates are formed solely from the distribution of runtimes from past examples of failure in similar systems [179].

One example of a type 1 prognostic technique is Weibull analysis [180]. In this method, the total runtime at failure is recorded over multiple similar systems. A Weibull distribution is then fitted to this data to approximate the distribution of these historical failure times. The Weibull distribution is calculated as in (5.1), where t is the failure time β is a shape parameter and λ is a scale parameter. This distribution is often used due to its flexibility in modelling different shapes of distribution [179].

$$p(t) = \frac{\beta}{\lambda}\left(\frac{t}{\lambda}\right)^{\beta-1} \tag{5.1}$$

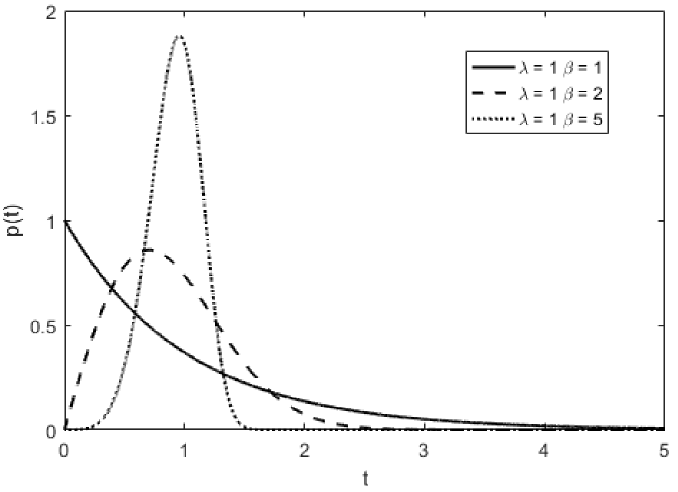


Figure 5.2 - Weibull distributions with varying shape parameter β values

These methods are clearly not suitable for new applications, where historical failure times of components will be sparsely available. In addition, time to failure analysis relies on the assumption that all systems have historically operated under similar conditions, and that failures will therefore progress at similar rates to those in the past [21]. This assumption is generally impractical in many industrial applications, as operating conditions can vary over time and vary between separate systems operating in the same environment (e.g. tidal arrays).

5.1.2 Type 2: Stressor-based

Stressor-based prognostic methods expand upon type 1 methods by incorporating operating conditions (environmental stressors) of the system into the prognostic model [179]. These stressors give an indication of the usage of components and may include loading, temperatures, vibrations, etc., modelling the failure rate of an average component operating under these measured conditions [21].

Examples of stressor-based techniques can include regression analysis, Markov chain models, Proportional Hazards Models, physics-of-failure models, and Life Consumption models [21].

The main assumption of type 2 techniques is that systems operating under the same conditions will fail at similar rates [21]. This allows for little variance between components operating in similar environments, but that may fail at different rates.

As with type 1 methods, type 2 prognostics techniques are also mostly dependent on historical examples of failure to construct accurate predictive models. This makes this category of technique also unsuitable for applying prognostics in new systems with limited past examples of failure.

5.1.3 Type 3: Degradation-based Prognostics

Finally, degradation-based prognostic methods predict the failure rates of specific components from their individual responses and operation [21]. This is performed through the measurement of a degradation parameter, which indicates the level of health of a component and is related to the probability of failure at that point in time [179]. This degradation parameter may be a parameter measured directly from the system or component, or may be a parameter generated through some operation that indirectly implies the level of damage.

This category of prognostics also requires knowledge of the degradation path, i.e. the trend or growth of a degradation parameter over time. This allows degradation parameters to be extrapolated forward in time to a failure threshold. Methods of modelling the degradation path broadly splits type 3 prognostic techniques into two categories: data-driven and model-based [181], where degradation is inferred either from historical data or from a model representing the underlying mechanisms of failure respectively.

This work finds type 3 prognostic techniques most suitable for new applications, in particular model-based methods. Using these methods, predictions are not solely reliant on the availability of historical data and can be inferred in part from the current state of individual components, based on current operating conditions, through the measurement of a degradation parameter. Models may also be constructed from knowledge of similar components operating in different environments, where such models can vary their degradation predictions to the current application based on operational measurements from the system, or by tuning parameters to best fit the observed trend of degradation.

The remainder of this chapter studies degradation-based prognostics in further detail, discussing how such methods can be implemented to give prognostic predictions within new applications where systems have limited operational histories.

5.2 Degradation-based Prognostics

The goal of degradation-based prognostics techniques is to estimate the future point in time k_{E} at which an estimated damage state of a component crosses some threshold T_{E} representative of failure. From here, the remaining useful life (RUL) of a component can be estimated as in (5.2). This process can be visualised as in Figure 5.3.

$$RUL = k_E - k \tag{5.2}$$

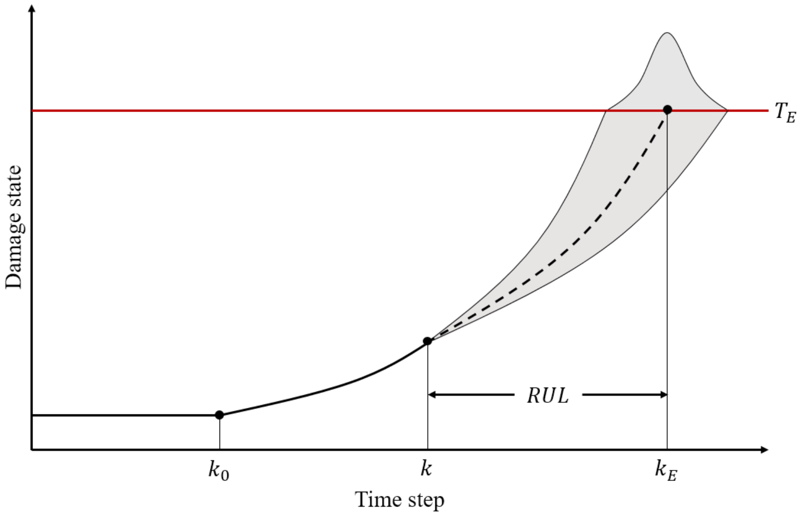


Figure 5.3 - Process of model-based prognostics through the extrapolation of an estimated damage state towards failure threshold

Degradation-based prognostic algorithms can be broadly separated into two categories: data driven and model-based [181]. A third category of hybrid methods can also be defined, where the results of both data driven and model-based methods are fused together, however this type of prognostics is not discussed in this work.

Data driven algorithms make predictions of the damage stage of a component based on data from past examples of failure [178]. This often involves using statistical or machine learning methods to fit a predictive model to past sensor data recorded during failure. Examples of such algorithms include, but are not limited to, Hidden Markov Modelling and neural networks [181].

A requirement of data driven approaches is the availability of run-to-failure data, where the quality of the dataset determines the performance of the chosen technique [181]. These methods are therefore not suitable for newly commissioned systems, where fault data will be extremely sparse.

Model-based prognostic algorithms combine knowledge of the physical degradation process with measured sensor data to make predictions [182]. In these techniques, knowledge of the physical degradation process is expressed through a parametric model, and is not solely dependent on the availability of run-to-failure data. This makes this category of degradation-based prognostics most suitable to new applications, where historical failure is often not available. Here, models can be developed based on knowledge of the degradation process, which can be inferred from physics-based knowledge of the specific system or from knowledge of the failure of similar components in different systems. Parameters of the model can then be tuned in operation to best fit the observed degradation trend of individual components.

5.2.1 Model Based Methods

Figure 5.4 [182] details the architecture of a model-based prognostics system. Here, the system converts inputs $\boldsymbol{u}_k$ (e.g. operational parameters) to observations or measurements $\boldsymbol{y}_k$ (e.g. condition parameters). Then a fault detection isolation and identification stage (e.g. anomaly detection and diagnostics) can be performed to identify a specific fault mode. Finally, prognostics is performed in two stages: damage estimation and prediction. Firstly, damage estimation is performed, where the distribution of current damage state $\boldsymbol{x}_k$ and degradation model parameters $\boldsymbol{\theta}_k$ are estimated from measurements $\boldsymbol{y}$. Secondly, prediction is performed, where the distribution of current damage state $\boldsymbol{x}_k$ is extrapolated forward in time based on the current estimate of model parameters $\boldsymbol{\theta}_k$, with hypothesised future inputs, to find the distribution of the RUL.

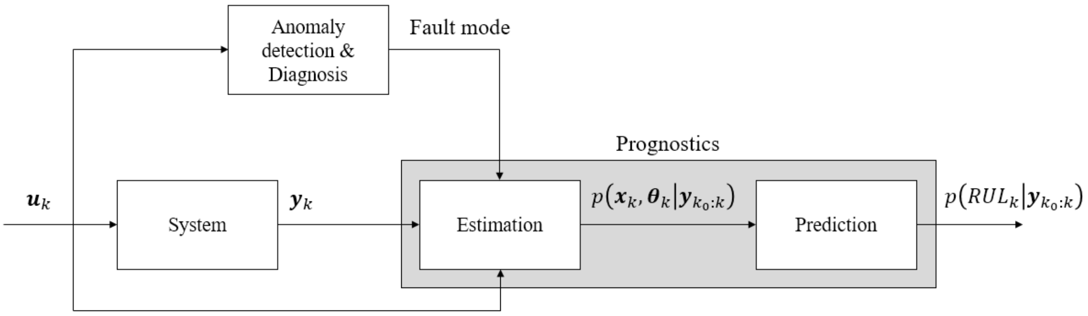


Figure 5.4 - Architecture of a model-based prognostics system, adapted from [182]

These distribution predictions can be achieved through Bayesian filtering approaches [181] where the uncertainty of the model and state estimates are modelled over time. In these approaches it is assumed that the current damage state $\boldsymbol{x}_k$ is dependent on the previous state $\boldsymbol{x}_{k-1}$ and operational conditions [181].

Bayesian filtering approaches include the Kalman filter, Extended Kalman filter, Unscented Kalman filter and particle filter. These methods are examined in more detail as follows.

5.2.1.1 Kalman Filter

The Kalman filter [183] makes the assumption that the state $\boldsymbol{x}$ at time step k can be inferred from the previous state at step $(k-1)$ as in (5.3), where $\mathbf{F}_k$ is a state transition model, $\mathbf{B}_k$ is a control-input model applied to inputs $\boldsymbol{u}_k$ (e.g. operational parameters) and $\boldsymbol{\omega}_k$ is process noise. In the Kalman filter, process noise is assumed to be normally distributed with a mean of zero and covariance $\mathbf{Q}_k$.

$$\boldsymbol{x}_k = \mathbf{F}_k \boldsymbol{x}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k + \boldsymbol{\omega}_k \tag{5.3}$$

Outputs or measurements of the true state $\boldsymbol{y}_k$ can be expressed as in (5.4), where $\mathbf{H}_k$ is an observation model and $\boldsymbol{v}_k$ is observation noise, also assumed to be normally distributed with a mean of zero and having covariance $\mathbf{R}_k$.

$$\boldsymbol{y}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k \tag{5.4}$$

From here, the Kalman filter is performed in two stages, prediction and estimation [184]. The first stage is the prediction stage, where the state estimate from the previous step in time $(k-1)$ is used to predict the state at the current point in time k. This is expressed in (5.5) and (5.6), where $\tilde{\boldsymbol{x}}$ donates state estimates and $\mathbf{P}$ donates the estimates' error covariance, a measure of the estimates' accuracy.

$$\tilde{\boldsymbol{x}}_{k|k-1} = \mathbf{F}_k \tilde{\boldsymbol{x}}_{k-1|k-1} + \mathbf{B}_{k-1} \boldsymbol{u}_{k-1} \tag{5.5}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^{\mathrm{T}} + \mathbf{Q}_k \tag{5.6}$$

The second stage is the update stage, where the current state estimate is combined with observations $\boldsymbol{y}_k$. This can be performed as in (5.7) to (5.11) [185], filtering $\tilde{\boldsymbol{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ to give final predictions $\tilde{\boldsymbol{x}}_{k|k}$ and $\mathbf{P}_{k|k}$, where $\mathbf{K}_k$ is a weighting known as the Kalman gain.

$$\tilde{\boldsymbol{y}}_k = \boldsymbol{y}_k - \mathbf{H}_k \tilde{\boldsymbol{x}}_{k|k-1} \tag{5.7}$$

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k \tag{5.8}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} \mathbf{S}_k^{-1} \tag{5.9}$$

$$\tilde{\boldsymbol{x}}_{k|k} = \tilde{\boldsymbol{x}}_{k|k-1} + \mathbf{K}_k \tilde{\boldsymbol{y}}_k \tag{5.10}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \tag{5.11}$$

Prognostics can then be performed by repeating the prediction stage in (5.5) and (5.6), iteratively estimating future damage states $\tilde{\boldsymbol{x}}_{k:k_E}$without measurement correction. This process can be stopped when the damage state exceeds a threshold T_E representative of component failure. The RUL of the damaged component can then be estimated as in (5.2).

5.2.1.2 Extended Kalman Filter

The original Kalman Filter is limited to linear systems. For nonlinear systems, or for non-Gaussian noise, there is no closed form solution to find the state space PDF [186]. However, the Kalman filter algorithm can be altered to solve such problems. One example is the Extended Kalman Filter (EKF), where nonlinear functions are linearized around the current estimate and the Kalman filter is then applied [187].

In this case, the state transition and observation models are allowed to be nonlinear, but must be differentiable [186]. These state transition and observation models can be expressed as in (5.12) and (5.13) respectively.

$$\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}, \boldsymbol{u}_k) + \boldsymbol{\omega}_k \tag{5.12}$$

$$\boldsymbol{y}_k = h(\boldsymbol{x}_k, \boldsymbol{v}_k) \tag{5.13}$$

Here, f and h cannot be directly applied to the covariance as in the Kalman filter. Instead, the Jacobian matrices (matrices of partial derivatives) **F** and **H** are calculated at each step k for f and h respectively with the predicted state at $(k-1)$, as in (5.14) and (5.15) [186].

$$\mathbf{F}_k = \left(\frac{\partial f}{\partial \boldsymbol{x}}\right) |\tilde{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{u}_k \tag{5.14}$$

$$\mathbf{H}_k = \left(\frac{\partial h}{\partial \boldsymbol{x}}\right) |\tilde{\boldsymbol{x}}_{k|k-1} \tag{5.15}$$

From here, the prediction and update steps of the Kalman filter are applied (as in (5.5) to (5.11)), where (5.5) becomes (5.16) to accommodate the nonlinear model.

$$\tilde{\boldsymbol{x}}_{k|k-1} = f\left(\tilde{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{u}_k\right) \tag{5.16}$$

5.2.1.3 Unscented Kalman Filter

Although the EKF provides a nonlinear solution for component damage prediction, it can give poor performance for systems where f and h are highly nonlinear, and in addition, requires the calculation of Jacobian matrices which can be computationally expensive for online applications [182].

The Unscented Kalman Filter (UKF) is an alternative approach and estimates the distribution of the state using a deterministic sampling technique called the unscented transform [188]. Weighted sample points (also known as sigma points) are sampled around the mean and propagated through the nonlinear functions to form an estimate of the new mean and covariance [182]. This can be expressed as in (5.17), where sigma points $\boldsymbol{\chi}^i$ can be passed through nonlinear function g to give new sigma points $\boldsymbol{\gamma}^i$. The new mean $\bar{y}$ and covariance $\mathbf{P}_{\boldsymbol{yy}}$ can then be calculated as in (5.18) and (5.19) respectively, where w^i is the weights of each i^{th} sigma point.

$$\boldsymbol{\gamma}^i = g(\boldsymbol{\chi}^i) \tag{5.17}$$

$$\bar{y} = \sum_i w^i \boldsymbol{\gamma}^i \tag{5.18}$$

$$\mathbf{P}_{\boldsymbol{yy}} = \sum_i w^i (\boldsymbol{\gamma}^i - \bar{y})(\boldsymbol{\gamma}^i - \bar{y})^T \tag{5.19}$$

The UKF then follows the same procedure as the Kalman filter and EKF, modifying the prediction and update stages to use sigma points. Including sigma points, the prediction stage becomes:

$$\widetilde{\boldsymbol{\chi}}^i_{k|k-1} = f\left(\widetilde{\boldsymbol{\chi}}^i_{k-1|k-1}, \boldsymbol{u}_{k-1}\right) \tag{5.20}$$

$$\widetilde{\boldsymbol{x}}_{k|k-1} = \sum_i w^i \, \widetilde{\boldsymbol{\chi}}^i_{k|k-1} \tag{5.21}$$

$$\widetilde{\boldsymbol{\gamma}}^i_{k|k-1} = h\left(\widetilde{\boldsymbol{\chi}}^i_{k|k-1}\right) \tag{5.22}$$

$$\widetilde{\boldsymbol{y}}_{k|k-1} = \sum_i w^i \, \widetilde{\boldsymbol{\gamma}}^i_{k|k-1} \tag{5.23}$$

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_k + \sum_i w^i \left(\boldsymbol{\chi}^i_{k|k-1} - \widetilde{\boldsymbol{x}}_{k|k-1}\right) \left(\chi^i_{k|k-1} - \widetilde{\boldsymbol{x}}_{k|k-1}\right)^T \tag{5.24}$$

The update stage then becomes:

$$\mathbf{P}_{\boldsymbol{yy}} = \mathbf{R} + \sum_i w^i \left(\widetilde{\boldsymbol{\gamma}}^i_{k|k-1} - \widetilde{\boldsymbol{y}}_{k|k-1}\right)\left(\widetilde{\boldsymbol{\gamma}}^i_{k|k-1} - \widetilde{\boldsymbol{y}}_{k|k-1}\right)^T \tag{5.25}$$

$$\mathbf{P}_{\boldsymbol{xy}} = \sum_i w^i \left(\boldsymbol{\chi}^i_{k|k-1} - \widetilde{\boldsymbol{x}}_{k|k-1}\right)\left(\widetilde{\boldsymbol{\gamma}}^i_{k|k-1} - \widetilde{\boldsymbol{y}}_{k|k-1}\right)^T \tag{5.26}$$

$$\mathbf{K}_k = \mathbf{P}_{\boldsymbol{xy}} \mathbf{P}_{\boldsymbol{yy}}^{-1} \tag{5.27}$$

$$\widetilde{\boldsymbol{x}}_{k|k} = \widetilde{\boldsymbol{x}}_{k|k-1} + \mathbf{K}_k \left(\boldsymbol{y}_k - \widetilde{\boldsymbol{y}}_{k|k-1}\right) \tag{5.28}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{\boldsymbol{yy}} \mathbf{K}_k^T \tag{5.29}$$

5.2.1.4 Particle Filter

Finally, the particle filter provides an alternative to Kalman filter based approaches and can be applied directly to nonlinear systems with non-Gaussian noise [185].

Using this method, the damage state of machinery is estimated by a set of values sampled from the state space (known as particles) weighted to represent the probability distribution. These particles can be expressed as in (5.30), where $x_k^{(i)}$ and $w_k^{(i)}$ are the state vector estimate and weight respectively for particle i at each time step k, for P number of particles [182].

$$\left\{x_k^{(i)}, w_k^{(i)}\right\}_{i=1}^{P} \tag{5.30}$$

The posterior density of these samples can be estimated as in (5.31), where $\delta_{\boldsymbol{x}_k^{(i)}}(d\boldsymbol{x}_k)$ is the Dirac delta function at each particle state estimate $\boldsymbol{x}_k^{(i)}$.

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) \approx \sum_{i=1}^{P} w_k^{(i)} \delta_{\boldsymbol{x}_k^{(i)}}(d\boldsymbol{x}_k) \tag{5.31}$$

Particle filtering can be performed through a number of different procedures. Sampling importance resampling (SIR) [189] is commonly used within prognostics [186]. In this procedure particles are resampled at each time step to stop degeneracy, a problem that causes single particles to dominate after very few iterations [185].

In this procedure, state estimates for each particle $\boldsymbol{x}_k^{(i)}$ are calculated by applying the state transition model (5.32) to each of the previous particle state estimates $\boldsymbol{x}_{k-1}^{(i)}$, giving (5.33), where $\boldsymbol{\omega}_{k-1}^{(i)}$ are generated from a process noise PDF.

$$\boldsymbol{x}_k^{(i)} \sim p\left(\boldsymbol{x}_k \middle| \boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{u}_{k-1}\right) \tag{5.32}$$

$$\boldsymbol{x}_k^{(i)} = f\left(\boldsymbol{x}_{k-1}^{(i)}, \boldsymbol{u}_{k-1}\right) + \boldsymbol{\omega}_{k-1}^{(i)} \tag{5.33}$$

Weights are then assigned to each particle based on $\boldsymbol{y}_k$ (5.34), calculated using the observation model as in (5.35), where $\boldsymbol{v}_k^{(i)}$ are generated from an observation noise PDF.

$$w_k^{(i)} = w_{k-1}^{(i)} p\left(\boldsymbol{y}_k \middle| \boldsymbol{x}_k^{(i)}\right) \tag{5.34}$$

$$\boldsymbol{y}_k = h\left(\boldsymbol{x}_k^{(i)}, \boldsymbol{v}_k^{(i)}\right) \tag{5.35}$$

Weights are then normalised as in (5.36) to give (5.37).

$$\widehat{\boldsymbol{w}}_k = \frac{\boldsymbol{w}_k}{\sum_{i=1}^{P} w_k^{(i)}} \tag{5.36}$$

$$\sum_{i=1}^{P} \widehat{w}_k^{(i)} = 1 \tag{5.37}$$

Finally, particles $\left\{x_k^{(i)}, w_k^{(i)}\right\}$ are resampled. This resampling is performed if a measure of the effective number of particles P_{eff} falls below some threshold P_T [185]. P_{eff} can be estimated based on the variance of particle weights as in (5.38). P number of samples are then taken from the current set of particles, with probabilities proportional to their weights.

$$\hat{P}_{\text{eff}} = \frac{1}{\sum_{i=1}^{P}\left(w_k^{(i)}\right)^2} \tag{5.38}$$

5.2.1.5 Random Walk Joint State-Parameter Estimation

In the previous methods, only the estimation of the damage state is described. However, in model based prognostics, the parameters of the model also must be estimated at each time step.

This can be achieved by augmenting the model parameters $\boldsymbol{\theta}$ to the state vector $\boldsymbol{x}$ [182]. Here, both the state estimate and model parameters will be filtered around observations $\boldsymbol{y}$. However, some form of evolution must also be added to each parameter over time.

One example of evolution is known as the random walk [182], where noise is added to each parameter after every time step. This is expressed as in (5.39), where ξ_{k-1} is a noise parameter sampled from a PDF of some noise distribution. This can be applied directly in the particle filter, however in Kalman filter approaches it is applied by extending the process noise matrix $\mathbf{Q}$.

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \xi_{k-1} \tag{5.39}$$

The variance of noise ξ effects the rate of convergence of parameter estimates as well as the estimation performance once parameter estimates converge. E.g. a large variance can give faster convergence but results in wider state estimates, whereas a small variance will give slow parameter convergence but will give tighter state estimates [182]. A number of methods exist to tune the variance of ξ, including approaches by M. Daigle and K. Goebel [190], and J. Lui and M. West [191].

5.3 Application of Model-based Prognostics in New Systems

Model-based prognostics has been identified as most suitable for applying prognostics in new applications where limited run to failure data exist. However, these methods require the selection of a suitable measure of degradation as well as a degradation model.

In this section, the process of identifying suitable degradation parameters and models is discussed. In this work, this has primarily focussed on model-based prognostics of gearboxes and bearings. From here a prognostics methodology is proposed, detailing how model-based prognostics can be performed with anomaly detection and diagnostics results.

5.3.1 Degradation Parameter

In model-based prognostics, a measure of degradation is used to estimate a damage state $\boldsymbol{x}$. This measured parameter is linked to observations of the system, $\boldsymbol{y}$.

For rotating machines, the degradation parameter can be linked to measured vibration sensor readings. Vibration was identified in chapters 3 and 4 as strongly related to the condition of rotating machine components. Chapter 3 linked the response of low resolution vibration measurements to operational parameters (generator rotation speed and output power) for anomaly detection. Chapter 4 then detailed how features from high resolution vibration signals can be used to make a diagnosis from faulty signals.

High resolution vibration features are commonly used as a degradation parameter for gearbox and bearing prognostics [192] [193] [194] [195] [196]. For example, degradation parameters may be taken from characteristic frequencies from the vibration signal. M. Gašperin, et al., [192] use the amplitude of the gear mesh frequency from a measured vibration signal to predict the RUL of a gear stage experiencing gear tooth wear through the particle filter.

However, vibration features used for prognosis are generally more widespread than individual frequencies and can instead measure larger bands of frequencies filtered from the original vibration signal. Y-X. Jia, et al., [193] measure the amplitude of a frequency band (between 3 kHz and 4.5 kHz) over time as a measure of degradation of a gearbox experiencing gear tooth wear. C. K. R. Lim and D. Mba [194] also use the amplitude of a frequency band (between 250 Hz and 2.5 kHz) to measure the degradation of a bearing experiencing an inner raceway fault.

In addition to the amplitude, other metrics can be used to measure degradation within specific frequency bands. For example, E. Sutrisno, et al., [195] measure degradation of a bearing using the kurtosis of a frequency band (between 5.5 and 6.0 kHz) over time. Y. Wang, et al., [196] show the kurtosis from select frequency bands is effective for prognostics of both bearing and gear tooth faults.

However, degradation parameters are not limited to high resolution vibration measurement. Methods have also been proposed for measuring the degradation of rotating machinery from low resolution

SCADA data [197] [198]. Here, low resolution signals can be used to build a model of normal behaviour (i.e. anomaly detection), where the residual from the model is used as a degradation parameter.

For example, S. Bulter, et al., [197] construct a model of low resolution temperature measurements from SCADA data, where the residual temperature between temperature readings and the model are used to measure the degradation of a wind turbine's main bearing. Alternatively, X. Jia, et al., [198] use SCADA data to build a model of a wind turbine's speed-power curve using principal component analysis and Gaussian mixture models, where decreasing levels of probability over time are used as a prognostic indicator.

5.3.2 Model Selection

A degradation model is used in model-based prognostics to estimate the progression of damage state $\boldsymbol{x}$ over time. These models are sometimes to referred to as physics-based models, as they approximate the physical mechanisms of failure over time [178].

Models used in different prognostic applications can vary from straightforward parametric functions, where parameters are tuned to best approximate degradation trends, to more complex models designed from first-principles of the system, where the underlying physical mechanics of failure are tracked over time using measurements from the system.

In many studies [199] [200] [192] [201] [195] [202] [194], more straightforward parametric functions are used in place of complex first-principle models. These models can provide a sufficient estimation of damage progression without a deep understanding of the underlying physics of failure, where the parameters of a function replace physical measurements or material properties. This modelling approach may be suitable for new systems where detailed physical knowledge of each component is not known, but models have been proven to effectively model the degradation of similar components in different environments.

The most straightforward degradation model is a linear model, where the damage state grows with a constant gradient over time. A function representing the linear growth of $\boldsymbol{x}_k$ over k steps in time is shown in (5.40) with parameters $\boldsymbol{\theta} = \{m, c\}$. This function can be rewritten in terms of the previous state $\boldsymbol{x}_{k-1}$ and adding process noise ω_k as in (5.41), where c is dropped making the parameter set become $\theta = m$ and where Δt is the difference in time between time steps $(k-1)$ and k as in (5.42).

$$\boldsymbol{x}_k = mk + c \tag{5.40}$$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + m\Delta t + \omega_k \tag{5.41}$$

$$t_k = t_{k-1} + \Delta t \tag{5.42}$$

A linear degradation model is used by M. Gašperin, et al., [192] to model the degradation of gear tooth wear over time. However, the linear model is found to be less accurate than a nonlinear population growth model. This population growth model is shown in (5.43) and (5.44), where the damage state is expressed as two vectors $\boldsymbol{x}_k = \{x_k^1, x_k^2\}$, the parameter set is $\boldsymbol{\theta} = \{a, b, c\}$ and $\boldsymbol{\omega_k} = \{\omega_k^1, \omega_k^2\}$ are process noise parameters. Here, this nonlinear model was able to converge to a more accurate RUL estimate sooner than a linear model.

$$x_k^1 = x_{k-1}^1 + a(b + cx_{k-1}^2)^2 + \omega_k^1 \tag{5.43}$$

$$x_k^2 = x_{k-1}^2 + \omega_k^2 \tag{5.44}$$

J. Sun, et al., [201] have proposed an alternative linear state model as in (5.45), where model parameters are $\boldsymbol{\theta} = \{\beta_0, \beta_1\}$ and ω_k is process noise. This is used to model engine wear, and is found to be more practical than a pure linear function in modelling faults that degrade suddenly or abruptly in operation.

$$\boldsymbol{x}_k = \beta_0 + \beta_1 \boldsymbol{x}_{k-1} + \omega_k \tag{5.45}$$

Exponential functions have also been used in some studies to model abrupt failure, as in [195] [199] and [200]. An exponential growth of damage state $\boldsymbol{x}_k$ over time steps k is shown in (5.46), with parameters $\boldsymbol{\theta} = \{a, b\}$. This can be rewritten in terms of the previous state $\boldsymbol{x}_{k-1}$ and adding process noise ω_k as in (5.47), where the parameter set is reduced to $\theta = b$ and Δt is the difference in time between time steps $(k-1)$ and k as in (5.42).

$$\boldsymbol{x}_k = ae^{bk} \tag{5.46}$$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} e^{b\Delta t} + \omega_k \tag{5.47}$$

Polynomial functions can also be used to model abrupt degradations [194] [202]. A polynomial growth of damage state $\boldsymbol{x}_k$ over time steps k is shown in (5.48), with parameter set $\boldsymbol{\theta} = \{a, b, c\}$. This can be rewritten in terms of the previous state $\boldsymbol{x}_{k-1}$ and adding process noise ω_k as in (5.49), where Δt is the difference in time between time steps $(k-1)$ and k as in (5.42) and the parameter set becomes $\boldsymbol{\theta} = \{m, \dot{m}\}$ where m is the derivative of $\boldsymbol{x}_k$ (the rate of change of the damage state) and $\dot{m}$ is the derivative of m. Polynomial functions have the advantage of being linear, and can therefore be used in Kalman filter methods, however may not be as accurate as exponential functions [194].

$$\boldsymbol{x}_k = ak^2 + bk + c \tag{5.48}$$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + m\Delta t + \dot{m}\frac{\Delta t^2}{2} + \omega_k \tag{5.49}$$

Although more straightforward parametric functions have been shown to adequately model degradation in many cases, they do not directly consider the usage of a system over time. More complex models designed from the first-principles of the system however, are intended for specific failures based on analysis of the underlying physical principles of the system. Such models output a degradation measure linked to the root cause of failure. Models can contain some tunable parameters themselves, but also take inputs $\boldsymbol{u}$ to describe the usage conditions (e.g. loading).

One example of such a model is the Paris model of crack growth [203]. This model is shown in (5.50), where damage state $\boldsymbol{x}_k$ is the crack size, loading inputs $\boldsymbol{u} = \{\Delta\sigma, N\}$ are stress range (measured in Pa) and number of cycles respectively, and model parameters are $\boldsymbol{\theta} = \{C, m\}$. This model is used by D. An, et al., [204] and F. Cadini, et al., [205] to model simulated crack growths in nuclear power applications.

$$\boldsymbol{x}_k = C\left(\Delta\sigma\sqrt{\pi\boldsymbol{x}_{k-1}}\right)^m dN + \boldsymbol{x}_{k-1} \tag{5.50}$$

A number of models have also been developed to model various types of gear tooth damage. F. Chaari, et al., [206] develop a model of gear tooth spalling based on meshing stiffness between two spur gears. This model is then used to identify the growth in amplitude of frequencies on a simulated vibration signal, which can be used as damage state estimate $\boldsymbol{x}_k$. This work is further expanded by X. Liang and M. J. Zuo [207] [208] developing a vibration model of crack growth within planetary gear stages based on gear mesh stiffness.

These models require detailed knowledge of the physical and material properties of components. This can often involve the use of techniques, such as finite element analysis [206], to simulate the response of these properties within a system over time. In addition, more complex models are also sensitive to inaccuracies or assumptions made during model development [21].

5.3.3 Approach for New Applications

Figure 5.5 expands upon the architecture of model-based prognostics given in Figure 5.4, p. 148, for new systems, integrating the process with previous anomaly detection and diagnostics condition monitoring stages.

The anomaly detection stage first indicates if a fault has occurred within the system (Chapter 3). If a fault is detected at an early stage it can be assumed that the fault may progress is severity over time before the component finally fails.

From here, degradation parameters can be examined. Suitable degradation parameters for rotating machine faults are discussed in section 5.3.1, p. 154. However, prognostics can only be accurately applied if parameters show a meaningful degradation trend. This can be assessed through the monotonicity metric, a measure of the underlying positive or negative trend of a degradation parameter [21]. This can be calculated as the difference between the fraction of positive and negative derivatives (5.51), where the derivative of a degradation parameter at time step k can be calculated as in (5.52), where Δt is the interval between each time step. Degradation parameters showing a suitable trend through the monotonicity measure can be taken to the next stage.

$$\text{monotonicity} = \left|\frac{\#\text{pos } d_k - \#\text{neg } d_k}{n-1}\right| \tag{5.51}$$

$$d_k = \frac{y_k - y_{k-1}}{\Delta t} \tag{5.52}$$

The model selection stage involves choosing a model to best describe the degradation path exhibited by the chosen degradation parameter. A list of candidate models should be available at this stage, consisting of parametric functions and/or physics-based models (as discussed in section 5.3.2, p. 155). At this stage, diagnostic information may help the selection of a suitable model, aided by engineering knowledge of the fault. For example, if a bearing fault is detected, an exponential model is more likely to describe the future degradation trend than a linear model [199].

If multiple models are suitable for the fault mode, the fit of candidate models can also be assessed through least squares fitting (discussed in detail in chapter 3, section 3.5.4.1, p. 52) where RMSE and/or R^2 values can be calculated for each model based on the currently available degradation data. Models returning the lowest RMSE and highest R^2 values indicate a best fit to the current data.

Finally, following the selection of a suitable degradation parameter and model, model-based prognostics can be performed through state estimation and prediction (as described in section 5.2, p. 147). Here, the particle filter was found to be the most appropriate method in this application, as different non-linear models can be easily interchanged, with different noise distributions.

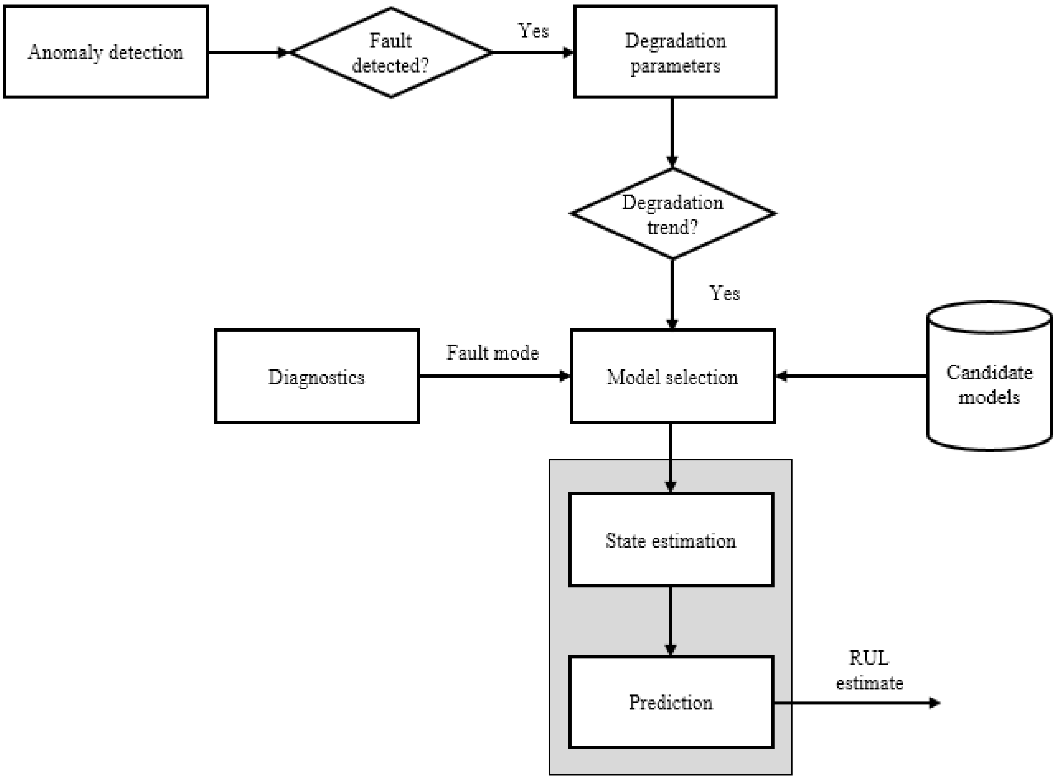


Figure 5.5 - Approach for model-based prognostics in new applications

5.3.4 Particle Filter Example

The methodology for applying prognostics in new applications was tested using data simulated to show the progression of a fault within the HS1000 gearbox, using the residual signal from the gearbox vibration model developed in chapter 3, section 3.6.1, p. 58. In this model, the normal response of the gearbox is defined by fitting a 3rd order Gaussian function to the trend of gearbox vibration against output power.

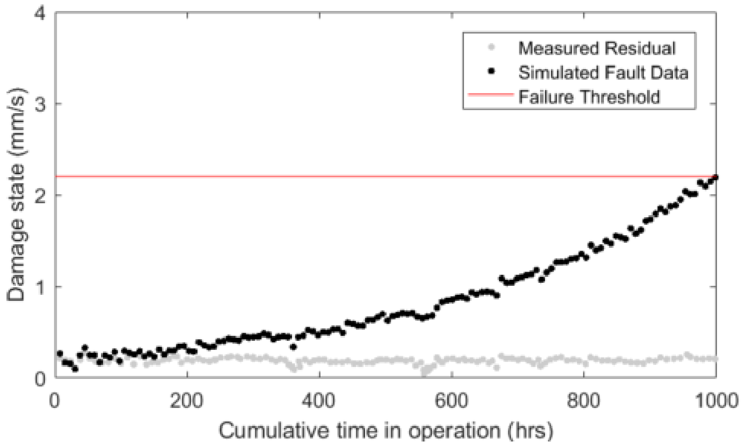
The particle filter was utilised in this case, due to its capability in using nonlinear degradation models and modelling non-Gaussian process noise.

5.3.4.1 Simulated Gearbox Degradation

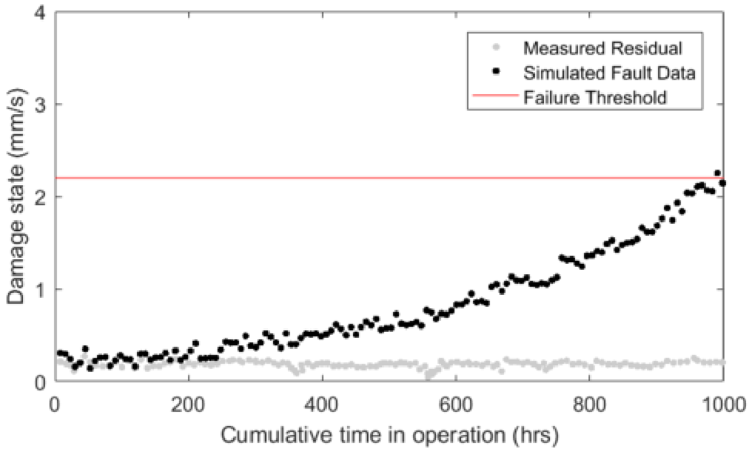
A simulated fault was introduced as in (5.53), adapting (3.41) from chapter 3, section 3.7.1.1, p. 70, where v_k is the raw measured vibration signal, and F_k is a fault signal with amplitude increased over time steps k through an exponential function with parameters $\boldsymbol{\theta} = \{a, b\}$. Here, the b parameter controls the rate of degradation. Gaussian noise is also added to this signal, with zero mean and variance of σ_ϵ. Figure 5.6 shows simulated degradation for a number of different b and σ_ϵ values.

$$y_k = v_k + (ae^{bk})F_k + \epsilon \tag{5.53}$$

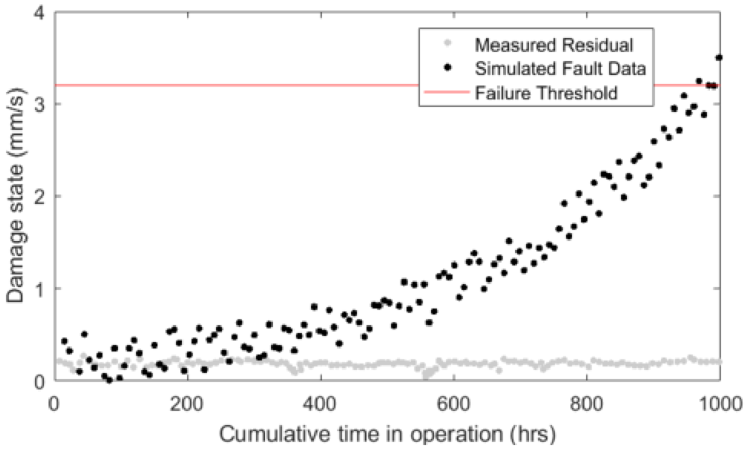
This exponential model of degradation does not take usage or operational parameters into account. This is somewhat accounted for by measuring k (each time step) as cumulative time in operation, where degradation only increases as the turbine is operational and rotating. This circumvents time periods during which the turbine is inactive (e.g. variable periods of inactivity between tidal cycles, and periods of planned or unplanned shutdowns). However, this assumes constant loading on the machine when operational. Improved prognostic models for tidal power may therefore take into account additional operational parameters (such as rotation speed, cumulative number of rotations, output power, cumulative power generated, or tidal flow rate) to model the rate of degradation as a function of the turbines loading.



(a) $a = 0.2$, $b = 0.0018$, $\sigma_{\epsilon} = 0.1$ and $T_E = 1.2$



(b) $a = 0.2$, $b = 0.0024$, $\sigma_{\epsilon} = 0.2$ and $T_E = 2.2$



(c) $a = 0.2$, $b = 0.0028$, $\sigma_{\epsilon} = 0.5$ and $T_E = 3.2$

Figure 5.6 - Simulated exponential degradation for varying b, η and T_E

5.3.4.2 Model Selection

In this example, model selection was performed through least squares regression (discussed in detail in chapter 3), where the parameters of a selection of candidate models are tuned to best fit the observed degradation trend.

In this study, no physics-based models of the tidal turbine component degradation were available. Therefore, candidate models included the parametric functions detailed in section 5.3.2, p. 155: linear function (5.41), population-growth function (5.43) and (5.44), alternative linear function (5.45), exponential function (5.47) and polynomial function (5.49). The RMSE between each fitted function and the observed degradation trend was calculated at each time step k, where the model returning the lowest RMSE was chosen for RUL prediction.

Table 5.1 shows the RMSE results for each model in each simulation with variable b and σ_η parameters at $k = 600$ hours. Here an exponential model can be seen to best fit the data, as expected, matching the exponential degradation model introduced in (5.53).

Table 5.1 - RMSE of different models at $k = 600$ hours

$\boldsymbol{b}$	$\boldsymbol{\sigma_\eta}$	**Linear RMSE**	**Pop. Growth RMSE**	**Alt. Linear RMSE**	**Exp. RMSE**	**Poly. RMSE**
0.018	0.1	0.0808	0.0624	0.0811	**0.0374**	0.0390
	0.2	0.0924	0.0748	0.0927	**0.0635**	0.0653
	0.5	0.1536	0.1520	0.1542	**0.1422**	0.1425
0.024	0.1	0.1741	0.0725	0.1748	**0.0391**	0.0530
	0.2	0.1731	0.0910	0.1738	**0.0673**	0.0735
	0.5	0.2354	0.1876	0.2363	**0.1501**	0.1587
0.028	0.1	0.2936	0.1543	0.2947	**0.0405**	0.0811
	0.2	0.3014	0.1890	0.3026	**0.0611**	0.0947
	0.5	0.3351	0.2030	0.3364	**0.1464**	0.1666

5.3.4.3 RUL Prediction

Following model selection, prognostics was performed through the particle filter. Predictions were made at each time step starting from $k = 600$ hours to the time of failure $k_E = 998$ hours.

First, the damage state $\boldsymbol{x}_k$ between the time of fault inception k_0 and the current time step k was estimated using Sampling Importance Resampling (SIR) (as in section 5.2.1.4, p. 152). An RUL prediction was then calculated by extrapolating the model from current time step k to k_E using the random walk (as in section 5.2.1.5, p. 153). This process is illustrated in Figure 5.7 for $k = 600$ hours. This process was performed iteratively for $k{:}\, k_E$, calculating an RUL prediction PDF at each time step.

The accuracy of each RUL prediction over time was measured through the cumulative relative accuracy (CRA) [209]. This accuracy measure varies between 0 and 1, where a value of 1 indicates a series of perfect RUL predictions over time. First, the relative accuracy RA can be calculated as in (5.54), where RA_k is the relative accuracy, a normalised measure of the difference between an RUL prediction ($\widehat{RUL}_k$) and the true RUL (RUL_k) at time point k. The CRA can then be calculated as in (5.55), the normalised sum of RA values between the starting prediction time k and the final time point at failure k_E.

$$RA_k = 1 - \frac{\left|RUL_{\mathrm{k}} - \widehat{RUL}_k\right|}{RUL_k} \tag{5.54}$$

$$CRA = \frac{1}{k_E - k - 1} \sum_{k}^{k_E} RA_k \tag{5.55}$$

Figure 5.8 shows the RUL prediction over time for degradation simulation where $b = 0.018$ and $\sigma_\epsilon = 0.2$. Here it can be seen that as k increases, the estimated RUL becomes more accurate and tighter around the true RUL.

The CRA values for a number of different simulation with varying b and η values are shown in Table 5.2. Increasing noise through σ_ϵ decreases the performance in each simulation. Increasing b, therefore increasing the rate of degradation through the exponential model, appears to increase the performance.

In each case, the particle filter achieved accuracies of above 0.8. This indicates the particle filter can provide sensible prognostic estimates even with varying levels of noise provided a suitable degradation model is available. However, it is important to note that this example uses a simplified fault model that assumes degradation is not directly linked to varying operating conditions. In practice, the varying usage of the turbine will have an effect the rate of degradation. Designing degradation models that incorporate usage parameters (such as tidal flow rate or generator rotation over time) is therefore suggested as future work.

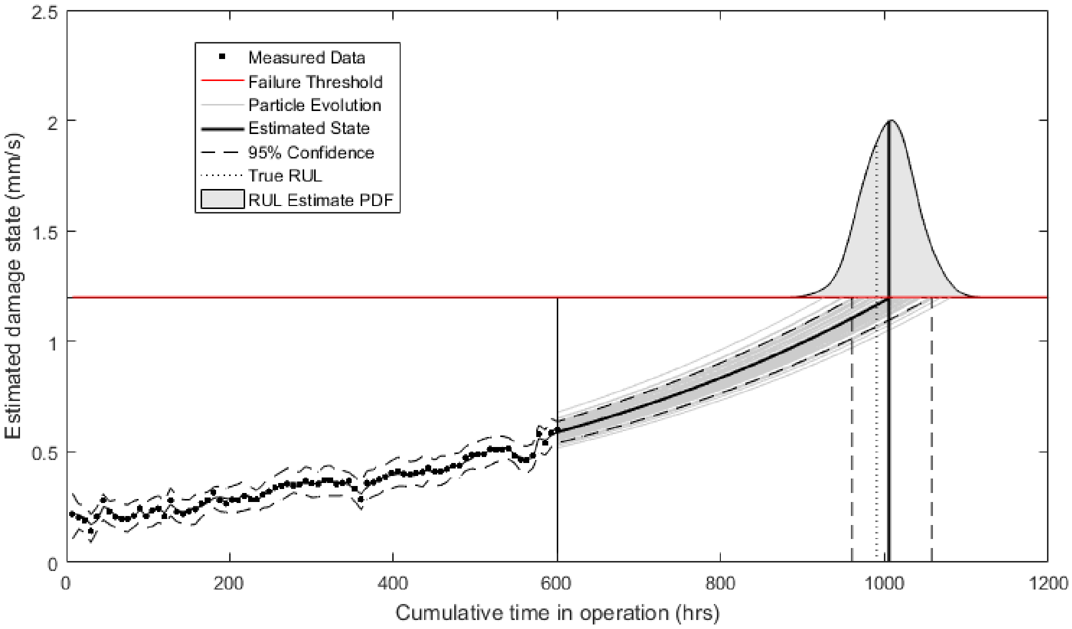


Figure 5.7 - Prediction of RUL after $k = 600$ hours for $b = 0.018$ and $\sigma_\epsilon = 0.2$

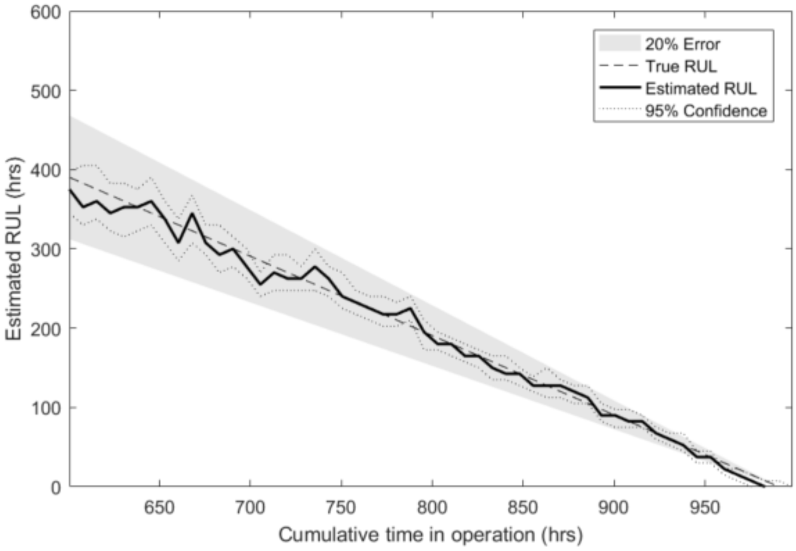


Figure 5.8 - RUL estimates over time for $b = 0.018$ and $\sigma_\epsilon = 0.2$

Table 5.2 - Cumulative Relative Accuracy (CRA) of RUL estimates for degradation simulations with varying growth b and noise variance σ_ϵ

$\boldsymbol{b}$	$\boldsymbol{\sigma_\epsilon}$	$\boldsymbol{CRA}$
0.018	0.1	0.873
	0.2	0.834
	0.5	0.804
0.024	0.1	0.928
	0.2	0.881
	0.5	0.803
0.028	0.1	0.943
	0.2	0.883
	0.5	0.881

5.4 Discussion

This chapter detailed prognostics methods for estimating the remaining useful life of equipment, focussing on techniques capable of making estimates for new systems with limited operating histories. Degradation-based prognostics techniques were found to be most suitable for new systems, where prognostic estimates are formed based on a measure of the degradation of specific components within a system.

In particular, model-based prognostics methods within this category were found to be most appropriate in comparison to data driven methods, where the degradation of a component over time is modelled through a degradation parameter, and extrapolated towards a failure threshold through a model that approximates the underlying degradation mechanics to give an RUL estimation. This method is the least reliant on historical failure data, as models can be inferred from the physics of the system or from knowledge of how similar components fail in different systems. The most suitable model-based technique was found to be the particle filter, a method capable of being applied to non-linear models with non-Gaussian noise.

An approach to applying model-based prognostics in new applications was then discussed, where considerations must be made in selecting an appropriate degradation parameter and degradation model for each component and fault mode.

With a focus on rotating machinery, suitable degradation parameters were found to include features from high resolution accelerometer data (measures of specific characteristic frequencies used in diagnostics or larger frequency bands), or residual signals from low resolution data model (i.e. anomaly detection models). Selection of an appropriate degradation parameter can then be made by observing which measures exhibit a suitable degradation trend once a fault has been detected.

Suitable degradation models included both physics-based models (where complex engineering knowledge of the system and its fault modes is used to design models of degradation) or more generic parametric functions (where the mechanics of degradation is approximated by tuning parameters). In each case, it is important to model both the estimated level of degradation and usage parameters to form accurate predictions. The predictability of periodic tidal flow rates (discussed in chapter 2, section 2.1, p. 5), may provide an advantage to the design of such prognostic models for tidal turbines, as future loading and operational parameters can be estimated more accurately than in other fields (e.g. wind power).

The CRISP-DM framework adapted for both anomaly detection and diagnostics can be further expanded to incorporate the prognostics approach for new applications (Figure 5.5, p. 159) as in Figure 5.9. Here, the data preparation stage can include the formation of degradation parameters (based on high resolution vibration data used for diagnostics, or residuals from anomaly detection models). If degradation measures show a suitable degradation trend, prognostic modelling can be performed, where diagnostic models can assist model selection. Evaluation of prognostic models can then be performed recursively as the fault progresses over time. Feedback to the prognostic modelling stage allows models or failure threshold values to be changed if measured degradation begins to differ from the chosen model or operators decide to alter acceptable degradation tolerances during operation respectively.

Finally, deployment will involve the estimation of the RUL of the component given a fault has been detected. This allows the operator to determine both the severity of the fault and the expected time until the severity crosses a pre-defined threshold. In having this information, operators can plan operational or maintenance activities more cost effectively. For example, the detection of a rapidly degrading fault can alert the operator to shut down machinery and schedule maintenance before the fault propagates to other components within the machine and causes costlier damage. Alternatively, if a component is estimated to be degrading at a slow rate the operation of the machine can be curtailed but not ceased and maintenance can be scheduled for a future date, minimising the costs due to machine downtime.

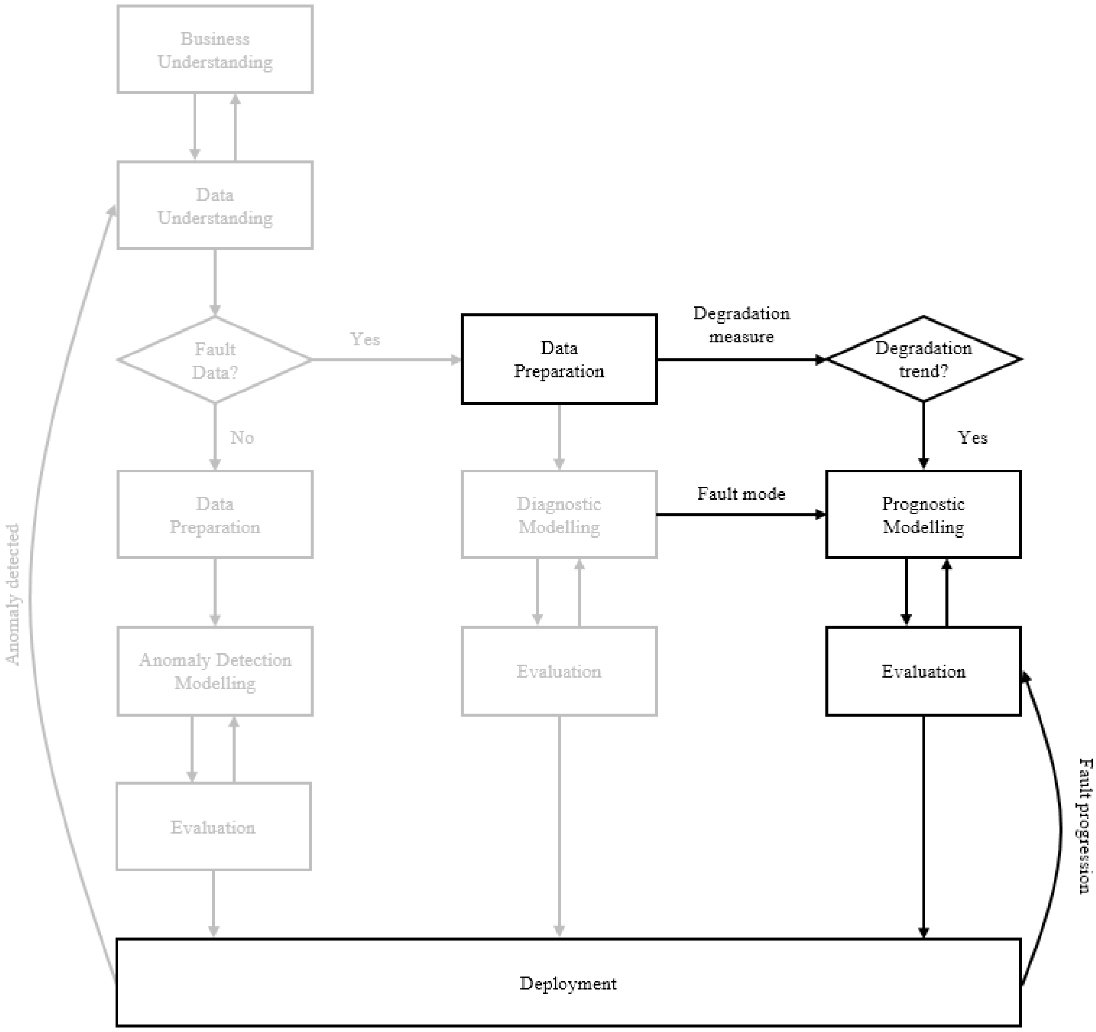


Figure 5.9 - CRISP-DM process expanded to include model-based prognostics

Chapter 6

6 Conclusions and Further Work

6.1 Summary and Conclusions

This thesis has presented an approach for applying condition monitoring in new applications, with a particular focus on tidal power.

In chapter 2, the need for a condition-based maintenance strategy in tidal power was discussed, reviewing the state of the art in turbine design and in condition monitoring. Condition monitoring was split into three categories, anomaly detection, diagnostics and prognostics, where each successive category requires more in depth knowledge of the system and its response to faults.

Chapter 3 detailed a data mining approach for anomaly detection, utilising CRISP-DM, a data mining framework, to define the normal response of the HS1000 tidal turbine from low resolution monitoring data. Here, data analysis found vibration to be the key condition parameter in describing the response of internal rotating components (the gearbox, high speed bearings and generator) against varying operating conditions. Generator rotation speed and output power were found key operational parameters in describing the response of the turbine. Using a technique known as change point analysis it was then shown that this data could be used to identify modes of the turbine's control scheme. This methodology can be used for defining the control modes of a system without prior knowledge of its design details or specifications.

From here, anomaly detection models for fault detection were then constructed using past operational data without faults present. It was found that reconstruction based anomaly detection models (such as regression or curve fitting methods) were best suited for gearbox vibration models, where data is likely to follow trends against operational parameters. Probabilistic modelling techniques (such as clustering algorithms or density models) were found to be most effective in modelling generator vibrations, where data groups into clusters dependent on control scheme modes. Finally, envelope modelling techniques were found to be the best modelling technique for low resolution bearing vibrations, where data amplitudes vary in relation to operational parameters but also continuously varies to zero.

The data mining process through CRISP-DM was proven to be an effective first stage for condition monitoring in new systems, allowing operators to construct models that learn from the normal response of machinery without past examples of faults and failure. However, while these models can detect changes in responses that may occur due to faults, the data still requires interpretation and analysis by operators to determine the cause of the anomaly.

Chapter 4 built upon the CRISP-DM model to include diagnostic modelling, where models can be constructed given past examples of faulty behaviour to automatically classify fault modes without further analysis by operators. Vibration data was again identified as a key indicator of faults for tidal turbines, where characteristic frequency responses of high resolution data can distinguish between fault types in rotating components. These characteristics for tidal turbine component failure modes were then detailed, before developing appropriate modelling techniques to perform automatic diagnoses in a condition monitoring system. Feature-based and deep learning classification approaches were then compared in this chapter, developing a deep learning approach through stacked autoencoder networks trained from spectrogram data that can perform fault diagnostics across many different failure modes from a single architecture. The use of this technique can provide accurate fault classifications without the need to develop tailored signal processing techniques to isolate each individual fault characteristic.

Finally, chapter 5 investigated prognostic methods, where the remaining useful life of components undergoing faults, degradation or aging can be estimated. This allows operators to track the progression of faults over time and identify how long a machine can be operated for until a shut down and maintenance is required. Degradation and model based prognostic methods were found to be most suitable for new applications where limited understanding of degradation trends or past run to failure data is available. An generic approach for implementing model based methods in new applications (such as tidal power) was then developed, considering the selection of suitable degradation parameters and models. Here, the particle filter was identified as the most appropriate prognostic technique, suitable for use with nonlinear models with different noise distributions. The original CRISP-DM methodology was finally further expanded to include this form of prognostic modelling, allowing operators to implement prognostic models with limited past examples of failure, and incorporating diagnostic information if available.

The final output of this thesis was a methodology for implementing anomaly detection, diagnostics and prognostics as part of a condition monitoring system for new applications where there is limited operational history, sparsely available historical data and a lack of knowledge or experience of how the machinery operates in practice. This methodology is based on an expansion of CRISP-DM, allowing operators to capture, learn from and model examples of faults or failure as they occur. The process was developed using experiences from data analysis of a newly commissioned tidal turbine in operation, however is generic and can be used across many different applications for condition monitoring or machine health management.

6.2 Main Contributions

The main contributions of this thesis were as follows:

- The identification of generator rotation speed and output power as key operational parameters and generator, gearbox and bearing vibrations as key condition parameters for condition monitoring of tidal turbines through the analysis of real tidal turbine data
- A data driven method for defining turbine control modes without prior knowledge of the system or its design details using change point analysis

- The development of anomaly detection models for fault detection in tidal turbine gearboxes, generators and bearings without using fault data
- A deep learning approach for the diagnosis of faults within rotating machinery from non-stationary vibration signals
- A methodology for implementing model-based prognostics in new systems with limited past examples of failure
- The expansion of CRISP-DM to provide a framework for implementing condition monitoring systems in new applications from limited data or knowledge of the machinery

6.3 Further Work

6.3.1 Additional Anomaly Detection Modelling

Following the CRISP-DM methodology discussed in chapter 3, models could be developed to define the response of additional components of the HS1000 beyond the gearbox, high speed bearing and generator.

For example, this could include:

- Modelling the response of the main bearing supporting the rotor and low speed shaft through temperature or vibration sensor data
- Modelling the response of external components, such as the blades, nacelle or support structure, through strain gauge measurements
- Modelling the electrical response of the generator through current and voltage measurements
- Modelling the response of additional turbine sub-systems such as lubricant or coolant pumps/motors through vibration, temperature or electrical sensor data.

6.3.2 Fault Simulation

In this work, gearbox vibration models were tested using data simulating a gear tooth fault and gear misalignment within the gearbox.

Additional fault simulations could be designed to replicate additional faults within the gearbox or other components (based on vibration responses given in chapter 4, section 4.1, p. 83) to further test the response of anomaly detection models.

In addition, fault simulations could be used to train fault diagnosis models, either through feature-based or deep learning classification approaches. Models trained on simulated faults could then be tested on real fault data as it becomes available.

6.3.3 Comparison of Data and Models between Separate Turbines

The work completed in this thesis involves modelling the response of the Andritz Hydro Hammerfest HS1000 tidal turbine, in operation at the EMEC in Orkney, Scotland. While this gave an accurate representation of the response of a commercial scale tidal turbine in real operation, it is unknown whether models developed in this work can be applied directly to an array of tidal turbines, operating in different areas and with alterations to their design specifications.

Future work could involve comparing data and models from the HS1000 turbine to different tidal turbines, detailing the effect of different operating environments and design specifications on the response of condition parameters. Here, the implications of transferring models developed on prototype devices to future tidal arrays can be better understood.

6.3.4 Development of Tidal-based Prognostic Models

In this work, model-based prognostics methods were identified as most suitable for new applications with limited past examples of component failure.

In these approaches, a mathematical model is used to estimate the degradation path of a component over time. This work identified suitable models used in other applications primarily for modelling degradation of bearings and gearboxes. However, none of these models have been adapted for usage with parameters specific to tidal power.

The predictability of tidal flow can offer a great advantage to the use of prognostics within this field, as future operating conditions can potentially be predicted accurately. Degradation models can therefore be developed to incorporate expected future loading conditions based on tidal predictions.

This future work could involve adapting more straightforward parametric functions to include loading parameters, or developing more complex models to track the progression of specific faults based on first principles analysis of the physics of failure.

References

[1] “Quantification of Exploitable Tidal Energy Resources in UK Waters,” ABP Marine Environmental Research Ltd., 2007.

[2] S. Benelghali, M. Benbouzid and J. F. Charpentier, “Marine Tidal Current Electric Power Generation Technology: State of the Art and Current Status,” in *IEEE Electrical Machines & Drives Conference*, Antalya, Turkey, 3-5 May 2007.

[3] A. I. Winter, “Differences in Fundamental Design Drivers for Wind and Tidal Turbines,” in *Proc. OCEANS 2011 IEEE*, Santander, Spain, Jun. 6-9, 2011.

[4] J. H. Williams, A. Davies and P. R. Drake, Condition-based maintenance and machine diagnostics, London, UK: Chapman & Hall, 1994.

[5] “Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009,” Official Journal of the European Union, 2009.

[6] “National Renewable Energy Action Plan for the United Kingdom: Article 4 of the Renewable Energy Directive,” Department of Energy & Climate Change, UK Government, 2009.

[7] “Third progress report on the promotion and use of energy from renewable sources for the United Kingdom,” Department of Energy & Climate Change, UK Government, 2016.

[8] “DUKES 2016 Chapter 6: Renewable sources of energy,” Department for Business, Energy & Industrial Strategy, National Statistics, 2016.

[9] H. H. H. Aly and M. E. El-Hawary, “State of the Art for Tidal Currents Electric Energy Resources,” in *Proc. 24th CCECE*, Niagra Falls, ON, May 8-11, 2011.

[10] P. Prickett, R. Grosvenor, C. Byrne, A. M. Jones, C. Morris, D. O'Doherty and T. O'Doherty, “Consideration of the Condition Based Maintenance of Marine Tidal Turbines,” in *9th EWTEC*, Southampton, UK, Sep. 5-9, 2011.

[11] P. Schureman, Manual of Harmonic Analysis and Prediction of Tides, U.S. Department of Commerce Coast and Geodetic Survey, 1958.

[12] M. A. Shields, L. J. Dillion, D. K. Woolf and A. T. Ford, “Stategic priorities for assessing ecological impacts of marine renewable energy devices in the Pentland Firth (Scotland, UK),” *Marine Policy*, vol. 33, no. 4, pp. 635-642, 2009.

[13] G. L. Mellor, “Chapter 11 Astronimical Tides,” in *Introduction to Physical Oceanography,* New York, NY, Springer-Verlag, 1996, pp. 196-212.

[14] J. J. King and T. Tryfonas, “Tidal Stream Power Technology - State of the Art,” in *Proc. OCEANS 2009 - EUROPE,* Bermen, Germany, May 11-14, 2009.

[15] J. Zhang, L. Moreau, M. Machmoum and P. E. Guillerm, “State of the Art in Tidal Current Energy Extracting Technologies,” in *First International Conference on Green Energy*, Sfax, Tunisia, 25-27 Mar. 2014.

[16] L. Chen and W.-H. Lam, “A review of survivability and remedial actions of tidal current turbines,” *Renewable and Sustainable Energy Reviews,* vol. 43, pp. 891-900, March, 2015.

[17] A. Uihlein and D. Magagna, “Wave and tidal current energy - A review of the current state of research beyond technology,” *Renewable and Sustainable Energy Reviews,* vol. 58, pp. 1070-1081, May, 2016.

[18] Y. Li and H. K. Florig, “Modeling the Operation and Maintenance Costs of a Large Scale Tidal Current Turbine Farm,” in *OCEANS 2006,* Singapore, 18-21 Sep. 2006.

[19] R. F. Orsagh, H. Lee, M. Watson, C. S. Byington and J. Powers, “Advanced Vibration Monitoring for Wind Turbine Health Management,” Impact Technologies, 2011.

[20] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe and G. Ekemb, “Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges,” *Energies,* vol. 7, pp. 2595-2630, 2014.

[21] J. Coble, “Merging Data Sources to Predict Remaining Useful Life - An Automated Method to Identify Prognostic Parameters,” PhD Thesis, Universty of Tennessee, Knoxville, TN, 2010.

[22] A. Grall, C. Bérenguer and L. Dieulle, “A condition-based maintenance policy for stochastically deteriorating systems,” *Reliability Engineering and System Safety,* vol. 76, pp. 167-180, 2002.

[23] P. P. Beaujean, T. M. Khoshgoftaar, J. C. Sloan, N. Xiros and D. Vendittis, “Monitoring Ocean Turbines: a Reliability Assessment,” in *16th ISSAT on Reliability and Quality in Design*, Washington, D.C., Aug. 5-7, 2010.

[24] S. Faulstich, B. Hahn and P. J. Tavner, “Wind tubrine downtime and its importance for offshore deployment,” *Wind Energy,* vol. 14, pp. 327-337, 2011.

[25] S. D. J. McArthur, C. D. Booth, J. R. McDonald and I. T. McFadyen, "An agent-based anomaly detection architecture for condition monitoring," *IEEE Trans. Power Systems,* vol. 20, no. 4, pp. 1675-1682, 2005.

[26] J. Z. Sikorska, M. Hodkiewicz and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing,* vol. 25, pp. 1803-1836, 2011.

[27] W. Yang, "18 - Condition monitoring of offshore wind turbines," in *Offshore Wind Farms*, Woodhead Publishing, 2016, pp. 543-572.

[28] W. Y. Liu, B. P. Tang, J. G. Han, X. N. Lu, N. N. Hu and Z. Z. He, "The structure healthy condition monitoring and fault diagnosis methods in wind tubrines: A review," *Renewable and Sustainable Energy Reviews,* vol. 44, pp. 466-472, 2015.

[29] P. Henríquez, J. B. Alonso, M. A. Ferrer and C. M. Travieso, "Review of Automatic Fault Diagnosis Systems Using Audio and Vibration Signals," *IEEE Trans. Systems, Man, and Cybernetics: Systems,* vol. 44, no. 5, pp. 642-652, 2014.

[30] P. Baraldi, F. Cadini, F. Mangili and E. Zio, "Model-based and data-driven prognostics under different available information," *Probabilistic Engineering Mechanics,* vol. 32, pp. 66-79, 2013.

[31] M. L. Wymore, J. E. Van Dam, H. Ceylan and D. Qiao, "A survey of health monitoring systems for wind turbines," *Renewable and Sustainable Energy Reviews,* vol. 52, pp. 976-990, 2015.

[32] H. D. Machado de Azevedo, A. M. Araujo and N. Bouchonneau, "A review of wind tubrine bearing condition monitoring: State of the art and challenges," *Renewable and Sustainable Energy Reviews,* vol. 56, pp. 368-379, 2016.

[33] P. Caselitz and J. Giebhardt, "Condition Monitoring and Fault Prediction for Marine Current Turbines," in *Proc. International Conference Ocean Energy*, Bremerhaven, Germany, Oct. 23-24. 2006.

[34] P. J. Tavner, "Review of condition monitoring of rotating electrical machines," *IET Electric Power Applications,* vol. 2, no. 4, pp. 215-247, 2008.

[35] R. Wald, T. M. Khoshgoftaar, P. Beaujean, J. C. Sloan and B. Raton, "A Review of Prognostics and Health Monitoring Techniques for Autonomous Ocean Systems," in *16th ISSAT on Reliability and Quality in Design*, Washington, D.C., Aug. 5-7, 2010.

[36] J. Duhaney, T. M. Khoshgoftaar, J. C. Sloan, B. Alhalabi and P. P. Beaujean, “A Dynamometer for an Ocean Turbine Prototype: Reliability through Automated Monitoring,” in *2011 IEEE 13th Int. Symp. HASE*, Boca Raton, FL, Nov. 10-12, 2011.

[37] R. Wald, T. M. Khoshgoftaar, P. Beaujean, J. C. Sloan and B. Raton, “Combining Wavelet and Fourier Transforms in Reliability Analysis of Ocean Systems,” in *16th ISSAT on Reliability and Quality in Design*, Washington, D.C., Aug. 5-7, 2010.

[38] J. Duhaney, T. M. Khoshgoftaar and R. Wald, “Applying Feature Selection to Short Time Wavelet Transformed Vibration Data for Reliability Analysis of an Ocean Turbine,” in *2012 11th ICMLA*, Boca Raton, FL, Dec. 12-15, 2012.

[39] J. Duhaney, T. M. Khoshgoftaar and J. C. Sloan, “Feature Selection on Dynamometer Data for Reliability Analysis,” in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, Boca Raton, FL, Nov. 7-9, 2011.

[40] J. Duhaney, T. M. Khoshgoftaar and A. Napolitano, “Studying the Effect of Class Imbalance in Ocean Turbine Fault Data on Reliable State Detection,” in *2012 11th ICMLA*, Boca Raton, FL, Dec. 12-15, 2012.

[41] J. Duhaney, T. M. Khoshgoftaar, I. Cardei, B. Alhalabi and J. C. Sloan, “Applications of Data Fusion in Monitoring Inaccessible Ocean Machinery,” in *16th ISSAT on Reliability and Quality in Design*, Washington, D.C., Aug. 5-7, 2010.

[42] J. Duhaney, T. M. Khoshgoftaar and J. C. Sloan, “Feature Level Sensor Fusion for Improved Fault Detection in MCM Systems for Ocean Turbines,” in *24th International FLAIRS Conference*, Palm Beach, FL, May 18-20, 2011.

[43] J. Duhaney and T. M. Khoshgoftaar, “Decision Level Fusion of Wavelet Features for Ocean Turbine State Detection,” in *2012 11th ICMLA*, Boca Raton, FL, Dec. 12-15, 2012.

[44] M. Mjit, P. P. Beaujean and D. J. Vendittis, “Smart Vibration Monitoring System for an Ocean Turbine,” in *2011 IEEE 13th International Symposium HASE*, Boca Raton, FL, Nov. 10-12, 2011.

[45] T. Burton, N. Jenkins, D. Sharpe and E. Bossanyi, Wind Energy Handbook, West Sussex, UK: John Wiley & Sons, 2011.

[46] R. Wirth, “CRISP-DM: Towards a Standard Process Model for Data Mining,” in *Proc. Fourth Internatioanl Conference on the Practical Application of Knowledge Discovery and Data Mining*, New York, NY, 2000.

[47] U. Fayyad and Piatetsky-Shapiro, “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine,* vol. 17, no. 3, pp. 37-54, 1996.

[48] D. L. Olson and D. Delen, Advanced Data Mining Techniques, Heidelberg, Germany: Springer, 2008.

[49] O. Maimon and L. Rokach, Data Mining and Knowledge Discovery Handbook, New York, NY: Springer, 2005.

[50] D. Hand, H. Mannila and P. Smyth, Principles of Data Mining, Cambridge, MA: MIT Press, 2001.

[51] J. L. Rodgers and W. A. Nicewander, “Thirteen Ways to Look at the Correlation Coefficient,” *The American Statistician,* vol. 42, no. 1, pp. 59-66, Feb, 1988.

[52] K. Pearson, “On lines and planes of closest fit to systems of points in space,” *Philosophical Magazine,* vol. 2, no. 11, pp. 559-572, 1901.

[53] H. Abdi and L. J. Williams, “Principal Component Analysis,” *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 2, no. 4, pp. 433-459, 2010.

[54] W. A. Taylor, “Change-Point Analysis: A Powerful New Tool for Detecting Changes,” Taylor Enterprises, Inc., 31 Dec. 2014. [Online]. Available: http://www.variation.com/cpa/tech/changepoint.html#Change-Point Analysis.

[55] S. Arlot and A. Celisse, “A survey of cross-validation procedures for model selection,” *Statistics Surveys,* vol. 4, pp. 40-79, 2010.

[56] M. A. F. Pimentel, D. A. Clifton, L. Clifton and L. Tarassenko, “A review of novelty detection,” *Signal Processing,* vol. 99, pp. 215-249, 2014.

[57] M. A. T. Figueiredo and A. K. Jain, “Unsupervised Learning of Finite Mixture Models,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 3, pp. 381-396, 2002.

[58] L. Rabiner and B. Juang, “An introduction to hidden Markov models,” *IEEE ASSP Magazine,* vol. 3, no. 1, pp. 4-16, 1986.

[59] M. Rosenblatt, “Remarks on Some Nonparametric Estimates of a Density Function,” *The Annals of Mathematical Statistics,* vol. 27, no. 3, pp. 832-837, 1956.

[60] G. J. McLachlan and K. E. Basford, Mixture Models: Inference and Applications to Clustering, New York, NY: marcel Dekker, 1988.

[61] D. P. Feliv, R. B. Chinnam, F. Tseng and P. Baruah, “An Industrial Strength Novelty Detection Framework for Autonomous Equipment Monitoring and Diagnostics,” *IEEE Trans. Industrial Informatics*, vol. 6, no. 4, pp. 767-779, 2010.

[62] Q. Chen, P. Goulding, D. Sandoz and R. Wynne, “The Application of Kernel Density Estimates to Condition Monitoring for Process Industries,” in *Proc. American Control Conference*, Philadelphia, PA, 26 Jun. 1998.

[63] A. J. Brown, V. M. Catterson, M. Fox, D. Long and S. D. J. McArthur, “Learning Models of Plant Behavior for Anomaly Detection and Condition Monitoring,” in *International Conference on Intelligent Systems Applications to Power Systems*, Kaohsiung, Taiwan, 2007.

[64] A. P. Dempster, N. M. Laird and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society, Series B,* vol. 39, no. 1, pp. 1-38, 1977.

[65] J. Blömer and K. Bujna, “Adaptive Seeding for Gaussian Mixture Models,” in *10th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Auckland, New Zealand, 19-22 Apr. 2016.

[66] M. C. Jones, J. S. Marron and S. J. Sheather, “A brief survey of bandwidth selection for density estimation,” *Journal of the American Statistical Association,* vol. 91, no. 433, pp. 401-407, 1996.

[67] B. W. Silverman, Density Estimation for Statistics and Data Analysis, London: Chapman and Hall, 1986.

[68] S. P. Lloyd, “Least Squares Quantization in PCM,” *IEEE Transactions on Information Theory,* vol. 28, no. 2, pp. 129-137, 1982.

[69] N. S. Altman, “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,” *The American Statistician,* vol. 46, no. 3, pp. 175-185, 1991.

[70] J. M. Peña, J. A. Lozano and P. Larrañaga, “An empirical comparison of four initialization methods for the K-Means algorithm,” *Pattern Recognition Letters,* vol. 20, no. 10, pp. 1027-1040, 1999.

[71] M. F. Lima, B. B. Zarpelao, L. D. H. Sampaioa, J. J. P. C. Rodrigues, T. Abrao and M. L. Proenca, “Anomaly detection using baseline and K-means clustering,” in *SoftCOM,* Bol, Croatia, 23-25 Sep. 2010.

[72] E. Kreyszig, Advanced Engineering Mathematics, New York, NY: John Wiley & Sons, 1979.

[73] P. Lancaster and K. Salkauskas, Curve and surface fitting: An introduction, London, UK: Academic Press, 1986.

[74] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning Internal Representations by Error Propagation,” in *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1*, vol. 323, Cambridge, MA, MIT Press, 1986, pp. 318-362.

[75] R. Bi, C. Zhou, D. M. Hepburn and J. Rong, “A NBM Based on P-N Relationship for DFIG Wind Turbine Fault Detection,” in *2015 International Conference on Smart Grid and Clean Energy Technologies*, Chengdu, China, 20-23 Oct. 2015.

[76] M. Schlechtingen and I. F. Santos, “Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection,” *Mechanical Systems and Signal processing,* vol. 25, pp. 1849-1875, 2011.

[77] J. Wolberg, Data Analysis Using the Method of Least Squares, Berlin: Springer, 2006.

[78] A. S. Goldberger, Econometric Theory, New York, NY: John Wiley & Sons, 1964.

[79] J. J. Moré and D. C. Sorensen, “Computing a Trust Region Step,” *SIAM Journal on Scientific and Statistical Computing,* vol. 3, pp. 553-572, 1983.

[80] D. W. Marquardt, “An Algorithm for Least-Squares Estimation of Nonlinear Parameters,” *Journal of the Society for Industrial and Applied Mathematics,* vol. 11, no. 2, pp. 431-441, 1963.

[81] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor and J. Platt, “Support vector method for novelty detection,” *Advances in Neural Information Processing Systems,* vol. 12, no. 3, pp. 582-588, 2000.

[82] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning,* vol. 20, pp. 273-297, 1995.

[83] J. Shawe-Taylor and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge, UK: Cambridge University Press, 2004.

[84] A. Purarjomandlangrudi, A. H. Ghapanchi and M. Esmalifalak, “A data mining approach for fault diagnosis: An application of anomaly detection algorithm,” *Measurement,* vol. 55, pp. 353-352, 2014.

[85] J. J. A. Costello, G. M. West, S. D. J. McArthur and G. Campbell, “Self-Tuning Routine Alarm Analysis of VIbration Signals in Steam Turbine Generators,” *IEEE Trans. Reliability,* vol. 61, no. 3, pp. 731-740, Sep. 2012.

[86] Z. I. Botev, J. F. Grotowski and D. P. Kroese, “Kernel density esimation via diffusion,” *Annals of Statistics,* vol. 38, no. 5, pp. 2916-2957, 2010.

[87] C. Scheffer and P. Girdhar, Practical Machinery Vibration Analysis & Predictive Maintenance, Burlington, MA: Newnes, 2004.

[88] O. D. Mohammed, M. Rantatalo, J. O. Aidanpää and U. Kumar, “Vibration signal analysis for gear fault diagnosis with various crack progression scenarios,” *Mechanical Systems and Signal Processing,* vol. 41, no. 1-2, pp. 176-195, Dec. 2013.

[89] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks,* vol. 61, pp. 85-117, 2015.

[90] A. Maszynska, “Vibration Diagnostics of Rotating Machinery Malfunctions,” *International Journal of Rotating Machinery,* vol. 1, no. 3-4, pp. 237-266, 1995.

[91] P. Tavner, L. Ran, J. Penman and H. Sedding, Condition Monitoring of Rotating Electrical Machines, London: The Institution of Engineering and Technology, 2008.

[92] P. Girdhar and C. Scheffer, Practical Machinery Vibration Analysis and Predictive Maintenance, Burlington, MA: Newnes, 2004.

[93] “Failure Analysis Gears-Shafts-Bearings-Seals,” Rexnord Industries, LLC, Gear Group, 1978.

[94] Y. Lei, J. Lin, M. J. Zuo and Z. He, “Condition mointoring and fault diagnosis of planetary gearboxes: A review,” *Measurement,* vol. 48, pp. 292-305, 2014.

[95] P. Eschmann, L. Hasbargen and K. Weigand, Ball and Roller Bearings, New York, NY: John Wiley, 1985.

[96] D. Felten, “Understanding Bearing Vibration Frequencies,” Electrical Apparatus Service Association, Inc., 2003.

[97] Lubrication Engineers, Inc., “How Bearings Fail,” *A Publication of the Lubrication Engineers Technical Department,* no. 96.

[98] Barden Precision Bearings, “Bearing Failure: Causes and Cures,” Schaeffler.

[99] B. P. Graney and K. Starry, “Rolling Element Bearing Analysis,” *Materials Evaluation,* vol. 70, no. 1, pp. 78-85, 2011.

[100] J. Pyrhönen, T. Jokinen and V. Hrabovcová, Design of Rotating Electrical Machines, West Sussex, UK: John Wiley & Sons, 2014.

[101] K. R. Fyfe and E. D. S. Munck, “Analysis of Computed Order Tracking,” *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187-205, 1997.

[102] H. Vold and J. Leuridan, “High Resolution Order Tracking at Extreme Slew Rates Using Kalman Tracking Filters,” *Shock and Vibration*, vol. 2, no. 6, pp. 507-515, 1995.

[103] S. L. Marple, “Computing the Discrete-Time "Analytic" Signal via FFT,” *IEEE Trans. Signal Processing*, vol. 47, no. 9, pp. 2600-2603, 1999.

[104] M. Lebold, K. McClintic, R. Campbell, C. Byington and K. 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*, Virginia Beach, VA, 1-4 May 2000.

[105] A. Aherwar and M. S. Khalid, “Vibration Analysis Techniques for Gearbox Diagnostic: A Review,” *International Journal of Advanced Engineering Technology*, vol. 3, no. 2, pp. 1-9, 2012.

[106] R. M. Stewart, “Some useful analysis techniques for gearbox diagnosis,” Institute of Sound and Vibration Research, Unversity of Southampton, 1977.

[107] J. J. Zakrajsek, “A Review of Transmission Diagnostics Research at NASA Lewis Research Center,” NASA and the US Army Research Laboratory, Cleveland, OH, 1994.

[108] H. R. Decker, R. F. Handschuh and J. J. Zakrajsek, “An enhancement to the NA4 gear vibration diagnostic parameter,” NASA and the US Army Research Laboratory, Cleveland, OH, 1994.

[109] H. R. Martin, “Statistical moment analysis as a means of surface damage detection,” in *Proceedings of the Seventh International Modal Analysis Conference*, Schenectady, NY, 1989.

[110] H. Yang, J. Mathew and L. Ma, “Vibration feature extraction techniques for fault diagnosis of rotating machinery: a literature survey,” in *Asia-Pacific Vibration Conference*, Gold Coast, Australia, 12-14 Nov. 2003.

[111] A. K. S. Jardine, D. Lin and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mechanical Systems and Signal Processing*, vol. 20, pp. 1483-1510, 2006.

[112] C. Rader and N. Brenner, “A new principle for fast Fourier transform,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 24, no. 3, pp. 264-266, 1976.

[113] N. Baydar and A. Ball, “Detection of Gear Deterioration Under Varying Load Conditions by Using the Instantaneous Power Spectrum,” *Mechanical Systems and Signal Processing,* vol. 14, no. 6, pp. 907-921, 2000.

[114] P. Tamilselvan, P. Wang, S. Sheng and J. M. Twomey, “A two-stage diagnosis framework for wind tubrine gearbox condition monitoring,” *International Journal of Prognostics and Health Management,* vol. 4, pp. 1-11, 2013.

[115] F. Al-Badour, L. Cheded and M. Sunar, “Non-Stationary Vibration Signal Analysis of Rotating Machinery via Time-Frequency and Wavelet Techniques,” in *10th International Conference on ISSPA*, Kuala Lumpur, Malaysia, 10-13 May 2010.

[116] J. B. Allen, “Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform,” *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol. 25, no. 3, pp. 235-238, 1977.

[117] M. C. Pan, P. Sas and H. van Brussel, “Nonstationary time-frequency analysis for machine condition monitoring,” in *TFTS' 96*, Paris, France, 18-21 Jun. 1996.

[118] N. Sawalhi and R. B. Randall, “Gear parameter identification in a wind turbine gearbox using vibraiton signals,” *Mechanical Systems and Signal Processing,* vol. 42, no. 1-2, pp. 368-376, 2014.

[119] M. S. Safizadeh, A. A. Lakis and M. Thomas, “Using Short-Time Fourier Transform in Machinery Diagnosis,” in *WSEAS*, Rio de Janeiro, 25-27 Apr. 2005.

[120] P. D. Samuel and D. J. Pines, “Vibration seperation methodology for planetary gear health monitoring,” *Proc. SPIE,* vol. 3985, pp. 250-260, 2000.

[121] M. Amarnath, S. Swarnamani and C. Sujatha, “Application of Wavelet Transform to Gearbox Fault Diagnosis,” in *12th National Conference on Machines and Mechanisms*, Guwahati, India, 16-17 Dec. 2005.

[122] D. Wang, Q. Miao and R. Kang, “Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition,” *Journal of Sound and Vibration,* vol. 324, no. 3-5, pp. 1141-1157, 2009.

[123] J. Rafiee, M. A. Rafiee and P. W. Tse, “Application of mother wavelet functions for automatic gear and bearing fault diagnosis,” *Expert Systems with Applications,* vol. 37, no. 6, pp. 4568-4579, 2010.

[124] X. Wang, V. Makis and M. Yang, “A wavelet apporach to fault diagnosis of a gearbox under varying load conditions,” *Journal of Sound and Vibration,* vol. 325, no. 3, pp. 1570-1585, 2010.

[125] R. B. Randal and J. Hee, “Cepstrum Analysis,” *Brüel & Kjær Technical Review,* vol. 3, pp. 3-40, 1981.

[126] M. Satyam, V. S. Rao and C. G. Devy, “Cepstrum Analysis - An Advanced Technique in Vibration Analysis of Defects in Rotating Machinery,” *Defence Science Journal,* vol. 44, no. 1, pp. 53-60, 1994.

[127] L. Nacib, K. M. Pekpe and S. Sakhara, “Detecting Gear Tooth Cracks using Cepstral Analysis in Gearbox of Helicopters,” *International Journal of Advances in Engineering & Technology,* vol. 5, no. 2, pp. 139-145, 2013.

[128] E. Alpaydin, Introduction to Machine Learning, Cambridge, MA: MIT Press, 2010.

[129] M. A. Aizerman, E. braverman, E. Braverman and L. Rozonoer, “Theoretical foundations of the potential function method in patter recognition learning,” *Automation and Remote Control,* vol. 25, pp. 821-837, 1964.

[130] B. E. Boser, I. M. Guyon and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*, Pittsburgh, PA, 27-29 Jul. 1992.

[131] M. Elangovan, V. Sugumaran, K. I. Ramachandran and S. Ravikumar, “Effect of SVM kernel functions on classification of vibration signals of a single point cutting tool,” *Expert Systems with Applications,* vol. 38, pp. 15202-15207, 2011.

[132] P. Konar and P. Chattopadhyay, “Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs),” *Applied Soft Computing,* vol. 11, pp. 4203-4211, 2011.

[133] V. Sugumaran and K. I. Ramachandran, “Effect of number of features on classification of roller bearing faults using SVM and PSVM,” *Expert Systems with Applications,* vol. 38, pp. 4088-4096, 2011.

[134] Z. Liu, H. Cao, Z. Chen, Z. He and Z. Shen, “Multi-fault classification based on wavelet SVM with PSO algorithm to analyse vibration signals from rolling element bearings,” *Neurocomputing,* vol. 99, pp. 399-410, 2013.

[135] L. Wenyi, W. Zhenfeng, H. Jiguang and W. Guangfeng, “Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM,” *Renewable Energy,* vol. 50, pp. 1-6, 2013.

[136] V. Muralidharan, V. Sugumaran and V. Indira, “Fault diagnosis of monoblock centrifugal pump using SVM,” *Engineering Science and Technology, an International Journal,* vol. 17, pp. 152-157, 2014.

[137] X. Liu, L. Bo and H. Luo, “Bearing faults diagnostics based on hybrid LS-SVM and EMD method,” *Measurement,* vol. 59, pp. 145-166, 2015.

[138] X. Zhang, Y. Liang, J. Zhou and Y. Zang, “A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM,” *Measurement,* vol. 69, pp. 164-179, 2015.

[139] S. Karamizadeh, S. M. Abdullah, M. Halimi, J. Shayan and M. J. Rajabi, “Advantage and drawback of support vector machine functionality,” in *2014 International Conference on Computer, Communications, and Control Technology,* Sabah, Malaysia, 2-4 Sep. 2014.

[140] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classificaiton Techniques,” *Informatica,* vol. 31, pp. 249-268, 2007.

[141] J. R. Quinlan, “Simplifying Decision Trees,” *International Journal of Man-Machine Studies,* vol. 27, no. 3, pp. 221-234, 1987.

[142] L. Breiman, J. H. Friedman, R. A. Olsen and C. J. Stone, Classificaiton and regression trees, Boca Raton, FL: CRC Press, 1984.

[143] B. S. Yang, C. H. Park and H. J. Kim, “An Efficient Method of Vibration Diagnostics for Rotating Machinery using a Decision Tree,” *International Journal of Rotating Machinery,* vol. 6, no. 1, pp. 19-27, 2000.

[144] H.-H. Lee, N.-T. Nguyen and J.-M. Kwon, “Bearing Diagnosis Using Time-Domain Features and Decision Tree,” *Advanced Intelligent Computing Theories and Applications With Aspects of Artificial Intelligence,* vol. 4682, pp. 952-960, 2007.

[145] N.-T. Nguyen, J.-M. Kwon and H.-H. Lee, “A Study on Machine Fault Diangosis using Decision Tree,” *Journal of Electrical Engineering & Technology,* vol. 2, no. 4, pp. 461-467, 2007.

[146] M. Boumahdi, J.-P. Dron and O. Cousinard, “On the extraction of rules in the identification of beairng defects in rotating machinery using decision tree,” *Expert Systems with Applications,* vol. 37, no. 8, pp. 5887-5894, 2010.

[147] I. Aydin, M. Karakose and E. Akin, “An apporach for automated fault diagnosis based on a fuzzy decision tree and boundary analysis of a reconstructed phase space,” *ISA Transactions,* vol. 53, no. 2, pp. 220-229, 2014.

[148] N. Gangadhar, H. Kumar, S. Narendranath and V. Sugumaran, “Fault Diagnosis of Single Point Putting Tool through Vibration Signal Using Decision Tree Algorithm,” *Procedia Materials Science*, vol. 5, pp. 1434-1441, 2014.

[149] A. N. Nandi, C. Liu and M. L. D. Wong, “Intelligent Vibration Signal Processing for Condition Monitoring,” in *Surveillance 7*, Chartres, France, 29-30 Oct. 2013.

[150] A. Kaushal and M. Shukla, “Comparative Analysis to Highligh Pros and Cons of Data Mining Techniques - Clustering, Neural Network and Decision Tree,” *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 651-656, 2014.

[151] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.

[152] B. Bagheri, H. Ahmadi and R. Labbafi, “Application of Data Mining and Feature Extraction on Intelligent Fault Diagnosis by Artificial Nearual Network and k-Nearest Neighbor,” in *XIX International Conference on Electrical Machines*, Rome, Italy, 6-8 Sep. 2010.

[153] J. Seshadrinath, B. Singh and B. K. Panigrahi, “Sensitive interturn fault diagnosis in induction machine using vibration analysis,” in *IEEE Fifth Power India Conference*, Haryana, India, 19-22 Dec. 2012.

[154] D. H. Pandya, S. H. Upadhyay and S. P. Harsha, “Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emmision data using APF-KNN,” *Expert Systems with Applications*, vol. 40, pp. 4137-4145, 2013.

[155] J. Tian, C. Morillo, M. H. Azarian and M. Pecht, “Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis,” *IEEE Trans. Industrial Electronics*, vol. 63, no. 3, pp. 1793-1803, 2016.

[156] P. Baraldi, F. Cannarile, F. Di Maio and E. Zio, “Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions,” *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 1-13, 2016.

[157] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, Cambridge, MA: MIT Press, 2016.

[158] H. Liu, C. Liu and Y. Huang, “Adaptive feature extraction using spase coding for machinery fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 558-574, 2011.

[159] S. Martin-del-Campo and F. Sandin, “Towards zero-configuration condition monitoring based on dictionary learning,” in *Proc. EUSIPCO*, Nice, France, 2015.

[160] V. M. Catterson and B. Sheng, “Deep Neural Networks for Understanding and Diagnosing Partial Discharge Data,” in *Proc. IEEE Electrical Insulation Conference,* Seattle, WA, 2015.

[161] L. Q. Qiao and L. J. Xun, “State of health estimation combining robust feature learning with support vector regression,” in *Proc. CCC,* Hangzhou, China, 2015.

[162] N. K. Verma, V. K. Gupta, M. Sharma and R. K. Sevakula, “Intelligent Condition Based Monitoring of Rotating Machines Using Sparse Coding,” in *Proc. IEEE Conference on Prognostics and Health Management*, Gaithersburg, MD, 2013.

[163] T. Junbo, L. Weining, T. Junfeng and W. Xueqian, “Fault Diagnosis Method Study in Roller Bearing Based on Wavelet Transform and Stacked Auto-encoder,” in *Proc. CCDC,* Qingdao, China, 2015.

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

[165] Y. LeCun, Y. Bengio and G. E. Hinton, “Deep Learning,” *Nature,* vol. 521, pp. 436-444, 2015.

[166] A. Ng, “CS294A lecture notes: Sparse autoencoder,” Stanford University, 2010.

[167] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning representations by back-propagating errors,” *Nature,* vol. 323, pp. 533-536, 1986.

[168] Y. LeCun, L. Bottou, G. B. Orr and K. R. Müller, “Efficient BackProp,” in *Neural Networks: Tricks of the Trade,* Springer Berlin Heidelberg, 2012, pp. 9-48.

[169] M. F. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” *Neural Networks,* vol. 6, no. 4, pp. 525-533, 1993.

[170] G. E. Hinton, S. Osindero and Y. W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation,* vol. 18, no. 7, pp. 1527-1554, 2006.

[171] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P. A. Manzagol, “Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion,” *Journal of Machine Learning,* vol. 11, pp. 3371-3408, 2010.

[172] Y. Bengio, “Learning Deep Architectures for AI,” *Foundations and Trends in Machine Learning,* vol. 2, pp. 1-127, 2009.

[173] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data With Neural Networks,” *Science,* vol. 313, no. 5786, pp. 504-507, 2006.

[174] B. A. Olshausen and D. J. Field, “Sparse Coding with an Overcomplete Bases Set: A Strategy Employed by V1?,” *Vision Research,* vol. 37, no. 23, pp. 3311-3325, 1997.

[175] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Annals of Mathematical Statistics,* vol. 22, no. 1, pp. 79-86, 1951.

[176] Y. Bengio, P. Lamblin, D. Popovici and H. Larachelle, “Greedy Layer-Wise Training of Deep Networks,” in *NIPS 2006,* Vancouver, Canada, 4-7 Dec. 2006.

[177] “PHM Challenge 2009 Dataset,” [Online]. Available: http://www.phmsociety.org/references/datasets.

[178] J. Coble, P. Ramuhalli, L. Bond, J. W. Hines and B. Upadhyaya, “A Review of Prognostics and Health Management Applications in Nuclear Power Plants,” *International Journal of Prognostics and Health Management,* vol. 6, pp. 1-22, 2015.

[179] W. J. Hines and A. Usynin, “Current Computational Trends in Equipment Prognostics,” *International Journal of Computational Intelligence Systems,* vol. 1, no. 1, pp. 94-102, 2008.

[180] R. B. Abernethy, The New Weibull Handbook, North Palm Beack, FL: Abernethy, 2004.

[181] J. I. Aizpurua and V. M. Catterson, “Towards a Methodology for Design of Prognostic Systems,” in *Annual Conference of the Prognostics and Health Management Society,* Coronado, CA, 18-24 Oct. 2015.

[182] M. Daigle, B. Saha and K. Goebel, “A Comparison of Filter-based Approaches for Model-based Prognostics,” in *IEEE Aerospace Conference 2012,* Big Sky, MT, 3-10 Mar. 2012.

[183] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Journal of Basic Engineering,* vol. 82, no. 35, pp. 35-45, 1960.

[184] R. Faragher, “Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation,” *IEEE Signal Processing Magazine,* vol. 29, no. 5, pp. 128-132, 2012.

[185] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Trans. Signal Processing,* vol. 50, no. 2, pp. 174-188, 2002.

[186] B. Saha, K. Goebel and J. Christophersen, “Comparison of prognostic algorithms for estimating remaining useful life of batteries,” *Trans. Institute of Measurement and Control,* vol. 31, no. 3-4, pp. 293-308, 2009.

[187] S. J. Julier and J. K. Uhlmann, “Unscented Filtering and Nonlinear Estimation,” *Proceedings of the IEEE,* vol. 92, no. 3, pp. 401-422, 2004.

[188] S. J. Julier and J. K. Uhlmann, “A new extension of the Kalman filter to nonlinear systems,” in *Proceedings of the 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls,* Orlando, FL, 1997.

[189] N. Gordon, D. Salmond and A. F. M. Smith, “Novel approach to non-linear and non-Gaussian Bayesian state estimation,” *IEE Proceedings F (Radar and Signal Processing),* vol. 140, no. 2, pp. 107-113, 1993.

[190] M. Daigle and K. Goebel, “Multiple damage progression paths in model-based prognostics,” in *IEEE Aerospace Conference,* Big Sky, MT, 5-12 Mar. 2011.

[191] J. Liu and M. West, “Combined Parameter and State Estimation in Simulation-Based Filtering,” *Sequential Monte Carlo Methods in Practice,* pp. 197-223, 2001.

[192] M. Gašperin, D. Juričić and P. Boškoski, “Prediction of the remaining useful life: An integrated framework for model estimation and failure prognostics,” in *IEEE Conference on Prognostics and Health Management,* Denver, CO, 18-21 Jun. 2012.

[193] Y.-X. Jia, L. Sun, G.-Y. Lin and W.-G. Wang, “Application of Rao-Blackwellized Particle Filtering for Estimating Remaining Useful Life of Gearbox,” in *International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering,* Sichuan, China, 15-18 Jul. 2013.

[194] C. K. R. Lim and D. Mba, “Switching Kalman filter for failure prognostic,” *Mechanical Systems and Signal Processing,* Vols. 52-53, pp. 426-435, 2015.

[195] E. Sutrisno, H. Oh, A. S. S. Vasan and M. Pecht, “Estimation of remaining useful life of ball bearing using data driven methodologies,” in *IEEE Conference on Prognostics and Health Management,* Denver, CO, 18-21 Jun. 2012.

[196] J. Wang, J. Xiang, R. Markert and M. Liang, “Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications,” *Mechanical Systems and Signal Processing,* Vols. 66-67, pp. 679-698, 2016.

[197] S. Butler, F. O'Connor, D. Farren and J. V. Ringwood, “A feasibility study into prognostics for the main bearing of a wind turbine,” in *IEEE International Conference on Control Applications*, Dubrovnik, Croatia, 3-5 Oct. 2012.

[198] X. Jia, C. Jin, M. Buzza, W. Wang and J. Lee, "Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves," *Renewable Energy,* vol. 99, pp. 1191-1201, 2016.

[199] N. Z. Gebraeel, M. A. Lawley, R. Li and J. K. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Transactions,* vol. 37, no. 6, pp. 543-557, 2005.

[200] A. Saxena, K. Goebel, D. Simon and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *International Conference on Prognostics and Health Management,* Denver, CO, 6-9 Oct. 2008.

[201] J. Sun, H. Zuo, W. Wang and M. G. Pecht, "Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance," *Mechanical Systems and Signal Processing,* vol. 28, pp. 585-596, 2012.

[202] L. C. K. Reuben, "Diagnostics and prognostics using switching Kalman filters," *Structural Health Monitoring,* vol. 13, no. 3, pp. 296-306, 2014.

[203] P. C. Paris and F. Erdogan, "A critical analysis of crack propagation laws," *Journal of basic engineering,* vol. 85, no. 4, pp. 528-533, 1963.

[204] D. An, J.-H. Choi and N. H. Kim, "Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab," *Reliability Engineering and System Safety,* vol. 115, pp. 161-169, 2013.

[205] F. Cadini, E. Zio and D. Avram, "Model-based Monte Carlo state estimation for condition-based component replacement," *Reliability Engineering and System Safety,* vol. 94, pp. 752-758, 2009.

[206] F. Chaari, W. Baccar, M. S. Abbes and M. Haddar, "Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of one-stage spur gear transmission," *European Journal of Mechanics A/Solids,* vol. 27, pp. 691-705, 2008.

[207] X. Liang and M. J. Zuo, "Dynamic Simulation of a Planetary Gear Set and Estimation of Fault Growth on the Sun Gear," in *International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering,* Sichaan, China, 15-18 Jul. 2013.

[208] X. Liang and M. J. Zuo, "Investigating vibration properties of a planetary gear set with a cracked tooth in a planet gear," in *Annual Conference of the Prognostics and Health Management Society*, Fort Worth, TX, 27 Sep. - 3 Oct. 2014.

[209] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha and M. Scwabacher, “Metrics for evaluating performance of prognostic techniques,” in *International Conference on Prognostics and Health Management*, Denver, CO, 6-9 Oct. 2008.