University of Cincinnati

Date: 8/14/2013

I, David Siegel, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Mechanical Engineering.

It is entitled:
Prognostics and Health Assessment of a Multi-Regime System using a Residual Clustering Health Monitoring Approach

Student's name: David Siegel

This work and its defense approved by:

Committee chair: Jay Lee, Ph.D.
Committee member: Canh Ly, Ph.D.
Committee member: Teik Lim, Ph.D.
Committee member: David Thompson, Ph.D.
Prognostics and Health Assessment of a Multi-Regime System using a Residual Clustering Health Monitoring Approach

A dissertation submitted to the Graduate School
Of the University of Cincinnati

in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
In the School of Dynamic Systems
Of the College of Engineering and Applied Science

by

David Siegel

August 2013

B.S. in Mechanical Engineering, University of Cincinnati (2007)
M.S. in Mechanical Engineering, University of Cincinnati (2009)

Committee Chair: Professor Jay Lee
Committee Member: Professor Teik Lim
Committee Member: Professor David Thompson
Committee Member: Dr. Canh Ly
ABSTRACT

Monitoring the health condition of machinery has been an area of research for quite some time. Despite several advancements, the application of conventional signal analysis and pattern recognition methods face several challenges when the operating variables such as load, speed, and temperature vary considerably for the monitored asset. The residual clustering approach addresses the multi-regime monitoring challenge by first modeling the baseline non-linear correlation relationship in the measured signal features and by providing predicted signal features. Calculating the residual signal features allows one to normalize the effect of the operating variables, since one is considering how the response of the system compares with the predicted response based on the baseline behavior. In many instances the degradation signature of a component or system is more pronounced under certain operating conditions. The clustering portion of the residual clustering method specifically addresses the regime dependent signature aspect and bases the health value on the monitoring regime in which the degradation signature is more prevalent.

This dissertation work highlights the mathematical framework and provides guidance on the appropriate processing methods for each portion of the approach. From simulation studies and wind speed data, the results highlight that the auto-associative neural network method provides the lowest prediction error when compared with regression, neural network, and principal component analysis methods. The results from this dissertation work also imply that the selection of the clustering algorithm does not significantly affect the calculated health value, and in general, most clustering algorithms appear suitable for detecting the problem using the residual clustering approach.
The feasibility of the residual clustering approach is demonstrated in three case studies. For the wind speed sensor health monitoring case study, the residual clustering method provides the most accurate health assessment of the wind speed sensors when compared with the other methods used by the 24 participants in the Prognostics and Health Management 2011 Data Challenge. The residual clustering approach also outperformed other multi-regime health monitoring methods such as a mixture distribution overlap method for the gearbox case study. The residual clustering method was also able to provide an early detection of a problem on the wind turbine rotor shaft with 26 days of advanced warning. The rotor shaft health value using the residual clustering approach had the most monotonic health trend when compared with three other multi-regime health monitoring methods for the wind turbine drivetrain case study.

The dissertation work shows that the residual clustering approach is fundamentally sound and should be considered along with the existing methods for multi-regime condition monitoring applications. The method appears to outperform many of the existing methods, and would be an appropriate monitoring algorithm if there is a nominal amount of correlation in the measured signals. Additional refinement of the approach can look into more sophisticated methods for threshold setting along with integrating a feature selection method into the residual clustering framework. In addition, algorithms for diagnosis and remaining useful life estimation for multi-regime condition monitoring applications would also require additional research and development work.
ACKNOWLEDGMENTS

I would like to express great appreciation to Professor Jay Lee for his guidance and support through my master’s and doctoral studies at the University of Cincinnati. I am also grateful to Professor Teik Lim, Professor Samuel Huang, Professor David Thompson, and Dr. Canh Ly for serving on my doctoral dissertation committee. I would like to express my gratitude to Mr. Patrick Brown for all the help he provided with proposals, budgets, administrative aspects, and marketing material that made working in the IMS Center an enjoyable experience. In addition, I appreciate the organizational and administrative skills of Michael Lyons, who helped organize customer meetings and several conferences, and allowed myself and other researchers to not worry about these administrative aspects.

I am also thankful to Dr. Mohamed AbuAli, Dr. Yixiang Huang, Dr. Xiangdong Zhu, Dr. Haixia Wang, and Dr. Masoud Ghaffari, for the guidance they provided me during their time as post-doctoral researchers at the IMS Center.

I also appreciate all the support and collaborative work I did with my colleagues during both my master’s and doctoral studies, including Dr. Edzel Lapira, Dr. Yan Chen, Mr. Hassan Al-Atat, Mr. Wenyu Zhao, Dr. Tianyi Wang, Mr. Viswesh Shauche, Dr. Lei Yang, Dr. Linxia Liao, Ms. Wenjing Jin, Mr. Muhammad Rezvani, Mr. Shanhun Yang, Mr. Hossein Davari Ardakani, Ms. Christina Lucas, Ms. Ann Kao, Ms. Xiaorui Tong, Mr. Mark McCrate, Mr. Tyler Skirtich, Mr. Feibai Zhu, Mr. Su Xu, Mr. Chaun Jiang, Mr. Behrad Bagheri, Mr. Zhe Shi, Mr. Chao Jin, Mr. Zongchang Liu, Mr. Eric Huang, Mr. Dustin Brisset, Mr. Mahmoud Hasan Pour Dargah, Mr. Ziyi Wang, Mr. Shuo Chang, Dr. Chein Chung Sun, Dr. Weihua Li, Mr. Jingliang Zhang, Mr. Fangji Wu, Mr. Shijin Wang, Mr. Bob Yu, Ms. Yang Yang, Ms. Meihui Wang, Ms. Fally Titikpina, and many others.
I also enjoyed my interaction with researchers from other IMS campuses including Dr. Seungchul Lee, Dr. Adam Brzezinski, Dr. Xiaoning Jin, Dr. Ahmad Almuhtady, and many others during various industrial advisory board meetings, joint proposals, or collaborative projects. Lastly, I would like to express thanks to my family for their constant support over the years.
TABLE OF CONTENTS

ABSTRACT ................................................................................................................................... II

ACKNOWLEDGMENTS .............................................................................................................. V

CHAPTER 1: INTRODUCTION ................................................................................................... 1

1.1 INTRODUCTION ...................................................................................................................... 1
1.2 MOTIVATION ........................................................................................................................... 2
1.3 RESEARCH OBJECTIVE .......................................................................................................... 3
1.4 CONTRIBUTIONS AND BROADER IMPACT ........................................................................... 4

CHAPTER 2: REVIEW OF RELATED LITERATURE ON PROGNOSTICS AND HEALTH
MANAGEMENT METHODS AND TECHNIQUES ........................................................................ 6

2.1 OVERVIEW OF DEVELOPING PREDICTIVE HEALTH MONITORING SYSTEMS ....... 6
2.2 DATA PRE-PROCESSING METHODS ...................................................................................... 10
2.3 OVERVIEW OF FEATURE EXTRACTION METHODS .............................................................. 14
2.4 WAVEFORM BASED FEATURE EXTRACTION ...................................................................... 16
2.4.1 TIME DOMAIN ANALYSIS ................................................................................................. 18
2.4.2 FREQUENCY DOMAIN ANALYSIS ...................................................................................... 20
2.4.3 TIME-FREQUENCY BASED METHODS .............................................................................. 22
2.4.4 ROLLING ELEMENT BEARING FEATURE EXTRACTION ALGORITHMS .................. 26
2.4.5 GEAR FEATURE EXTRACTION ALGORITHMS ................................................................. 29
2.5 FEATURE SELECTION METHODS ......................................................................................... 36
2.6 HEALTH ASSESSMENT METHODS ....................................................................................... 38
2.7 HEALTH DIAGNOSIS METHODS .......................................................................................... 41
2.8 REMAINING USEFUL LIFE PREDICTION METHODS ............................................................ 42

CHAPTER 3: PRIOR METHODS USED FOR THE PROGNOSTICS AND HEALTH
ASSESSMENT OF A MULTI-REGIME SYSTEM ..................................................................... 45

3.1 INTRODUCTION ...................................................................................................................... 45
3.2 LOCAL MODEL APPROACH ................................................................................................. 47
3.3 NORMALIZATION METHODS ................................................................................................. 49
1.1 Introduction

The maintenance philosophy in industry has been evolving from a reactive based maintenance approach to a more predictive condition based maintenance approach. However, in order to perform maintenance based on the condition of an asset, improvement in several areas is needed; such as better sensors, more robust health estimation algorithms, and more reliable failure prediction methods. Despite several advances in this field, certain aspects of technology are not currently mature enough to meet the demands for some of the more challenging applications. In particular, when a system is operating under dynamic and changing operating regimes, accurate health estimation and remaining useful life prediction becomes a more difficult task and many of the conventional health estimation techniques do not suffice.

Although robust and accurate remaining useful life estimation provides the most actionable information for performing condition based maintenance, the focus of this study is on improved methods for health assessment. The rationale is that many of the failure prediction methods use the information from health estimation algorithms as initial conditions or inputs into the prediction routine. Therefore, achieving reliable remaining useful life prediction algorithms would require improved health assessment models. Although an accurate similarity based prediction algorithm was presented by Wang et al. in 2009 [50] for systems operating under multiple regimes, the method required a substantial number of run-to-failure training sets which are difficult to obtain in practice. The present study focuses on the more common and challenging situation in which only baseline data from the system is provided. Furthermore, for
systems operating under multiple regimes, it is quite common that the degradation is more noticeable in certain operating conditions compared to others. Conventional approaches that use local models in each regime or methods that compare the feature distributions for the entire operating space might provide unsatisfactory results.

This research study proposes the development of a residual clustering health assessment framework to specifically address the previously mentioned challenges in health estimation for systems operating in multiple operating conditions. This introductory chapter will highlight the motivation for this research, the research objectives, and some broader impacts that can result from this work. In addition, the later chapters will review the prior work, present the mathematical framework for the method, and evaluate the feasibility of the method with three condition monitoring case studies.

1.2 Motivation

Although the subject of monitoring the health condition of machinery or assets has been considered for quite some time, many of the conventional approaches are based on the assumption of constant operating conditions, such as constant load or speed. There is substantial prior work on the area of condition monitoring of machinery for assets that operate in a single operating regime, which include algorithms related to signal analysis, signal fusion, pattern recognition, and failure prediction. However, some of the more challenging monitoring applications deal with assets and machines that operate under multiple loads, temperatures, and speeds, in which these operating parameters vary considerably during the assets usage. Example monitoring applications in which this multi-regime situation exists includes wind turbines, excavators used in mining, rotorcraft drivetrain components, machine tools, and aircraft engines. These examples span industrial sectors that include manufacturing, automotive, mining, power generation, and aerospace. There is an apparent need across many industrial sectors for analysis methods and algorithms to monitor the health condition of a
system operating in multiple regimes. Considering this recent need, there is some prior work on
developing analysis methods that are suited for monitoring multi-regime systems. A review of
the prior work in multi-regime health monitoring is presented in Chapter 2. However, the survey
of the prior work suggested that there are still many undressed challenges for multi-regime
health monitoring. These include the following:

- The condition monitoring signals and features are highly influenced by the operating
  conditions, and thus improved methods are needed to normalize the effect of the
  operating variables so one can deduce that the observed changes are only due to
  component or system degradation.

- The degradation or fault signature is normally more pronounced under a subset of the
  operating regimes and thus the prior art that uses distribution overlap methods or local
  modeling approaches do not fully account for this aspect.

1.3 Research Objective

The objective is to develop an effective health monitoring technique for a multi-regime
system in which the degradation signature is likely more pronounced under a subset of the
operating conditions. In this scenario, the challenges include the following: a lack of a priori
knowledge of which operating regime is most conducive for detecting the signature, the
measured signals are significantly affected by the operating conditions, and only data from a
baseline state is provided. In order to formally address these challenges, the proposed
research will have an in depth study on the residual clustering methodology in order to:

- Establish a methodology and framework for the key procedures in this approach.
• Find the best practices in the residual clustering approach, including the residual processing algorithm, the clustering algorithm and parameters, and the figure of merit health calculation.

• Benchmark the residual clustering monitoring method with other multiple regime health monitoring methods and evaluate the methods in various case studies.

1.4 Contributions and Broader Impact

The research conducted in this study provides a novel approach for health estimation of a multi-regime system using auto-associative models and a clustering based health calculation. Health estimation is a fundamental aspect of the prognostics and health management field. In addition, systems that operate in multiple regimes represents a difficult class of problems for prognostics and health monitoring in which conventional approaches do not provide satisfactory results. The research work is evaluated with three case studies in which the results are quite promising. Scientist and engineers that are researching and developing prognostic and health management (PHM) solutions can build upon the residual clustering approach for some of the more challenging multi-regime monitoring applications. The research presented in this study provides a meaningful contribution to scientist and engineers in the PHM community and it is expected that further research will be conducted that either enhances the current residual clustering approach or benchmarks this approach with newly developed methods.

The outcome of this research could potentially affect different sectors of the economy, including the energy sector and manufacturing. Many manufacturing equipment such as machine tools and industrial robots, operate under different speed and loading conditions and more accurate and robust health monitoring methods are needed for these situations. An accurate multi-regime health monitoring system could be used to take proactive maintenance on manufacturing equipment to prevent costly unexpected breakdowns and increase productivity.
The multi-regime health monitoring approach is also amendable for monitoring wind turbines. Improvement in wind turbine monitoring systems could enhance their reliability and make them a more cost competitive form of power generation. There are other multi-regime monitor applications for aircraft, ground vehicles, and mining equipment that also have similar needs for improvement maintenance and reliability that could benefit from the proposed multi-regime health monitoring approach.
CHAPTER 2: REVIEW OF RELATED LITERATURE ON PROGNOSTICS AND
HEALTH MANAGEMENT METHODS AND TECHNIQUES

The review of related literature on the field of prognostics and health management is presented in order to provide the reader an overall perspective on the current state of the art. The overview of the different prognostic and health management algorithms and how to select the appropriate methods was initially presented by the author in 2013, at the Machine Failure Prevention Technology (MFTP) conference [1]. This review builds upon that conference papers discussion of PHM algorithms by providing a more thorough review of each method. In addition, much of the material for the vibration based signal processing and feature extraction methods are based on the author’s previous comparative study for vibration condition monitoring, which is published in the Wind Energy journal [2].

2.1 Overview of Developing Predictive Health Monitoring Systems

The development of a prognostics and health monitoring system consist of a series of steps that are outlined in the flow chart provided in Figure 1. An initial understanding of the system is imperative prior to developing a health monitoring system. Engineering and domain knowledge of the system is a prerequisite for determining what existing signals are available, the potential sensors that can be added to the system, and the failure modes the component or system can experience. In addition, it would be advantageous to have some prior knowledge on the motivation and potential cost savings the monitoring system can provide; the cost related to unplanned downtime or failures can influence the selection of the data acquisition hardware and onboard computing resources for the health monitoring and prognostic algorithm. The selection
of the appropriate data acquisition hardware and which signals to monitor are key next steps in the development of the health monitoring system. This also includes determining the appropriate sampling rates for each signal, the interval or triggering for collecting data, where to store the raw or processed data, among other factors.

Figure 1: Prognostics and Health Management Development Flow Chart
For complicated monitoring systems with multiple sensors, the signal validation step becomes an important step. It was noted in a recent study on algorithms for automatic data validation of vibration signals, that the wind turbine vibration monitoring system evaluated in the case study was burdened with numerous data acquisition errors [3]. Intuitively, one can imagine that without a data validation step, all the subsequent feature extraction or more advanced algorithmic processing of the data would not result in useful health information. In addition to signal validation, the data pre-processing tools can include removing outlier data samples, extraction context information such as the operating conditions, and additional processing algorithms for checking the quality of the extracted features.

The selection of the appropriate feature extraction algorithms is usually application dependent, in that it depends on the type of the signals that are collected. Established methods and techniques are available for rotating machinery using vibration signals as well as other method signals such as motor current. For low frequency sampled signals such as temperature, pressure, or controller signals, various residual based processing methods or statistical features are quite appropriate. A more detailed discussion on the appropriate feature extraction techniques are provided in Section 2.3.

After developing the feature extraction algorithms for the given application, usually a sample set of data is needed to evaluate the proposed methods; this can consist of seeded fault data from a smaller scale test-bed or an initial set of data from the monitored fleet. At this stage in the development process, various data quality metrics related to signal quality, and also the quality of the feature with respect to the monitoring task such as clusterability [4] or trendability can be evaluated [5]. If the data quality appears to be at a suitable level, the subsequent steps in the health monitoring system development can proceed. However, if their appears to be a substantial amount of signal validation problems or the extracted features appear of low quality, further refinement in the data acquisition step and feature extraction algorithm are needed.
If there is historical data from multiple health states, the use of a feature selection method can provide a more robust health model by including the most discriminate set of features. Various techniques for feature selection are discussed in more detail in Section 2.6, however, the use of engineering knowledge and expert selection can be used if there is not sufficient data for performing feature selection. The fusion of multiple features into a single health metric can be used to detect whether the system is in an anomalous condition as well as to provide a single indicator for tracking the system or component degradation. The selection of the appropriate health assessment algorithm can be based on the available data for training the model as well as the complexity required for the monitoring task. In many applications, only data from a baseline condition is available for training the health model; thus distance from normal methods or hypothesis testing become more attractive methods compared to regression or neural network based methods.

Additional development could be required if diagnostics is a required functionality for the monitoring system. Various machine learning and expert knowledge algorithms for diagnostics are discussed in more detail in Section 2.10. In addition, if the monitoring system is required to provide an estimation of the remaining useful life, this would require selecting an appropriate prediction algorithm. Although in many instances the prognostic algorithms are less validated than the health assessment or diagnostic algorithms, various approaches and algorithms are available and are discussed further in Section 2.11. Accurate data driven prognostic models typically require extensive training data sets for learning the appropriate degradation models [6]. Stochastic filtering methods require a dynamics model to describe the physical degradation process which are difficult to obtain and usually only apply to a certain subset of failure modes of the monitored system or component [7].

The last set of steps in the health monitoring and prognostic development process includes an evaluation step. At this stage, initial data from the fleet or seeded fault data from a test-bed
can be used to evaluate the performance of the health monitoring and prognostic method. Various well established criteria exist for evaluating anomaly detection and diagnostic algorithm results. Receiver operating curves and confusion matrices are common ways of presenting and analyzing the results for diagnostic algorithms [8]. In addition, newly developed metrics for remaining useful life prediction algorithms are also available. The prediction metrics not only account for the prediction accuracy but also consider the time horizon in which the prediction is performed [9]. Based on the monitoring requirements for the given application, a decision can be made on whether any of the additional processing modules need adjustment or refinement. If no adjustments are required, the final monitoring algorithm can be deployed to the fleet of assets.

2.2 Data Pre-processing Methods

A listing of a set of data pre-processing tools along with the merits and disadvantages of each method are provided in Table 1. In many applications, one or more of these methods are needed for improving the data quality and ensuring that the data is suitable for additional processing and algorithm development. Signal validation is particular important for ensuring there are no sensor or data acquisition errors. However, developing a signal validation algorithm requires some prior knowledge about the signal characteristics or the signal distribution. In addition, many of the signal validation metrics require a threshold for determining whether the signal is valid and should be included for additional analysis [3]. For systems operating in multiple operating conditions, the data acquisition would typically be triggered to collect data under certain operating conditions or at predefined intervals. For more complicated systems, it might be necessary to have more advanced logic or methods for identifying the operating regime.
### Table 1: Summary of Data Pre-Processing Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Signal Validation</td>
<td>Requires some prior knowledge but effective in particular for vibration signal validation.</td>
<td>Thresholds are needed when using a signal validation metric.</td>
</tr>
<tr>
<td>2 Regime Identification</td>
<td>Context information is important for developing baseline data sets in each operating condition.</td>
<td>More sophisticated methods are needed for identifying the operating regime if the system changes operating conditions quickly.</td>
</tr>
<tr>
<td>3 Outlier Removal</td>
<td>Outlier can dramatically increase the false alarm rate.</td>
<td>Inclusion of domain knowledge is important for improving the accuracy of the outlier removal step.</td>
</tr>
<tr>
<td>4 Feature Clustering Quality Check</td>
<td>Clustering tendency metrics can be used to test whether classification algorithms would be suitable.</td>
<td>Applicable only to detection and diagnosis problems.</td>
</tr>
<tr>
<td>5 Feature Trendability or Monotonic Check</td>
<td>Important to check whether any of the features have a consistent trend, are monotonic, and have a similar failure value.</td>
<td>More suited for the task of prognostics and remaining useful life evaluation.</td>
</tr>
</tbody>
</table>
Further development of the health monitoring system would likely need information on the operating conditions. An example would be a rotating machinery application in which it is important to know the operating load and rotational speed since the vibration features are influenced by those operating variables [10]. The regime information allows one to develop baseline data sets in each operating condition and have a fair comparison and local health model in each operating condition.

Various methods exist for removing outlier instances from the signal or extracted features; minimum spanning tree methods or density based methods such as local outlier factor can be used [11]. However, these methods are purely based on the data distribution or characteristics. Domain knowledge could aid the outlier removal algorithm with the inclusion of rules and logic based on engineering knowledge of the measured signals and system. Additional data pre-processing tools include recently developed data quality metrics specifically designed for evaluating the extracted features used for developing prognostics and health management systems. Spectral clustering methods were used to evaluate the clustering tendency of the extracted features, in which the evaluation determined the number of clusters and the distribution of the individual clusters [4]. The intuition of this method is that features extracted under different health states should have well defined and separable clusters. By evaluating the clustering tendency, one can examine the feature quality and whether the extracted features are suitable for discriminating between the different health conditions. In addition, specific metrics for evaluating whether the features have a monotonic trend, consistent failure values, and repeatable degradation patterns, are also available for evaluating whether the feature set is suitable for prognostics [5]. It would be advantageous to perform the initial evaluation on whether the features are suitable for prognostics prior to spending significant resources on developing health and prediction models.
It is perhaps best to illustrate the data pre-processing methods with an example condition monitoring application. In this example, accelerometer data was collected from a spindle bearing test-rig in which the bearing ultimately locked up and failed at the end of the test. In addition to applying a variety of vibration based feature extraction methods, a signal validation check was also performed. Based on the prior work in [3], it was suggested to see how many consecutive values were the same in a vibration signal. If the vibration signal is measuring only noise due to the system not running or has cabling or other issues, there would be substantial more instances of consecutive samples having the same value. This consecutive sample check was performed on each vibration waveform that was collected.

The results of this signal validation method are presented in Figure 2, with the right most plots showing the signal validation results and the left most plots showing two example vibration

![Figure 2: Example Vibration Run to Failure Feature Data and Signal Validation Results](image-url)

The results of this signal validation method are presented in Figure 2, with the right most plots showing the signal validation results and the left most plots showing two example vibration
feature trends. The consecutive sample check on the vibration waveform shows that normally a vibration waveform would only have a few instances of this occurring; however there are two data files when this occurs over 90 times. These two files were collected when the motor was stopped and the system was not running. The vibration signal validation method correctly identified and removed these files, and further processing could be applied to the remaining data files. The left-most plots illustrate that these two files also had abnormally low feature values (highlighted in green). Although this situation only occurred twice, performing this signal validation could be quite important in an industrial application when these problems could occur more frequently. Having a signal validation step that can remove these outlier occurrences can provide a better trend in the features and thus provide more robust inputs for the health assessment and prediction models.

2.3 Overview of Feature Extraction Methods

Numerous methods and algorithms are available for extracting characteristics or features from the measured signals; an overview of some of the available techniques are provided in Table 2. For high frequency type signals such as vibration or current, there are well established signal processing and feature extraction methods for extracting information from the time and frequency domain representation of the signal [12]. For rolling element bearings, mechanical shafts, and gear wheels, there are several specific processing methods for extracting degradation features for these components [13]. Advanced signal processing methods for separating the vibration sources, and demodulation methods for extracting the amplitude and phase modulation characteristics are particular useful for extracting features for gear and rolling element bearing components [14]. In addition, there are several established motor current signature analysis techniques for detecting motor or generator related defects [15]. Although it is advantageous to use the component specific feature extraction methods for the high
frequency vibration and current signals, they usually require a higher sampling rate, more computation, and more costly data acquisition systems.

Table 2: Overview of Feature Extraction Techniques

<table>
<thead>
<tr>
<th>Method or Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 High Frequency Based Feature Extraction Methods</td>
<td>Frequency domain, envelope, and time-frequency features are very appropriate for highly sampled signals such as vibration and current signals.</td>
<td>Requires a higher sampling rate and more costly data acquisition.</td>
</tr>
<tr>
<td>2 Residual Based</td>
<td>More suited for low frequency signals and signals with a potential correlation, residuals can be calculated without significant domain knowledge.</td>
<td>Can involve training a neural network which requires more computation and potential for over fitting.</td>
</tr>
<tr>
<td>3 Statistics for each process segment or time slice</td>
<td>Ideal for process signals and provides a simple way of capturing the key aspects of the measured signal.</td>
<td>Requires context information for identifying the various time slices of a process signal.</td>
</tr>
<tr>
<td>4. Time Statistics</td>
<td>Requires the least amount of domain knowledge or context information, and also is simple to implement.</td>
<td>Provides less specific information than statistics for each process segment.</td>
</tr>
</tbody>
</table>
For applications in which the monitored set of signals consists of low frequency signals such as temperature, pressure, and other controller signals, a different set of feature extraction algorithms would be recommended. Residual based processing algorithms, based on regression, auto-associative neural networks, or principal component based methods would be appropriate method for correlated low frequency signals [16]. For these types of algorithms, the baseline representation of the system is used for calculating a predicted sensor value which can be compared to the measured signal for generating residuals. The potential drawback of these methods is that they could involve training a neural network, which requires more computation and can result in over fitting. The extraction of various statistical parameters is a straightforward but effective approach for characterizing the system condition from the available process or controller signals. In many instances, more insight can be gain by extracting statistics during different time slices. A time slice could represent a different motion or action that is being performed by the monitored system [17]. If the context information regarding the process signals is not available, then extracting time statistics without any segmentation is a suitable alternative.

2.4 Waveform Based Feature Extraction

For high frequency dynamic signals such as vibration, motor current, acoustic emission, or acoustic sound data, there is a wealth of information that can be extracted from the measured waveform. Signal transforms, along with advanced filtering and data processing method are crucial signal processing steps prior to extracting relevant features or condition indicators from the measure signals. Section 2.4.1 to Section 2.4.5 provides a discussion of some of the more commonly used signal transform methods along with a review of the more established condition indicators for gear and bearing components.

However, prior to discussing in detail those signal processing and feature extraction methods, it is best to present a typical measured waveform signal and discuss the qualitative
characteristics that can be used to assess the condition of the monitored component. An example vibration waveform measured from a two stage parallel gearbox is presented in Figure 3, in which the top plot is from the baseline condition and the bottom plot is for the eccentric idler gear condition. From a visual observation, one can observe higher peak to peak vibration amplitude levels for the eccentric gear condition when compared with the baseline signal. However, this higher amplitude situation could arise from other problems, including shaft imbalance or misalignment, other gear defects such as pitting on the gear teeth, and even from rolling element bearing defects.

Figure 3: Gearbox Vibration Waveform - Top Plot (Baseline), Bottom Plot (Eccentric Gear)

Perhaps more indicative of the eccentric gear signature is the pattern and periodicity of the higher vibration signature. If one were to further examine this waveform, they would notice that
the eccentric gear signature pattern is causing an amplitude modulation effect in which the modulation frequency is the idler shaft speed and the carrier frequency is the gear mesh frequency for that gear pair. This example highlights the qualitative signal characteristics, but various tools and mathematics are needed to extract this information in a consistent and robust manner. The specific method for extracting amplitude modulation gear condition features is reviewed in Section 2.4.5 and also discussed in more detail with the gearbox case study in Chapter 6.

2.4.1 Time Domain Analysis

The time domain analysis is based on extracting statistics from the measured signal waveform. These time domain statistics consists of amplitude information, such as the peak to peak value of the waveform, the root mean square (RMS), crest factor, among others. In addition, characteristics of the signal distribution can be extracted including higher order statistical moments such as kurtosis. In practice, the time domain statistical features are more suited for the task of anomaly detection or health assessment and less ideal for providing detailed diagnosis and identification of the particular degraded component or failure mode that is occurring. Intuitively, this would make sense, since higher amplitude vibration or more impacts could be attributed to a multitude of different failure modes, from shaft imbalance, to a cracked gear tooth, to a spall on an inner race of a rolling element bearing.

The utility of extracting features from the raw time waveform for anomaly detection are only modest at best, and can usually be improved by applying filtering methods which enhance the fault signature. One of the more promising filtering methods is a technique called spectral kurtosis, which is aimed at finding the optimal frequency band for recovering that impulsive fault signature that could be hidden in the raw vibration waveform. A brief review of the calculation procedure and example results are provided, and the interested reader is referred to the work by Antoni et al. [18] and Combet et al. [19] for a more detailed discussion on the use of spectral
kurtosis for filtering vibration signals. The initial step in this algorithm is to calculate the short time Fourier transform of the vibration signal, denoted by $H(t,f)$. Equation 1 indicates that the average value of the fourth power of $H(t,f)$ is divided by the mean square value of $H(t,f)$, which provides a kurtosis value as a function of frequency. The Wiener filter is constructed using the kurtosis values for each frequency bin as shown in Equation 2; the frequency bin is only included if the kurtosis value is above a statistical threshold at a given confidence level [18]. The Wiener filter is then multiplied by the frequency domain representation of the original signal, $X(f)$, and the result is transformed back to the time domain as indicated in Equation 3. The advantages of this method is that the signal is filtered without any a priori knowledge of which frequency band to filter in, and instead is based on which frequency band is most impulsive.

$$K_r(f) = \frac{\langle H^4(t,f) \rangle}{\langle H^2(t,f) \rangle^2} - 2$$

(1)

$$\hat{W}(f) = \begin{cases} \sqrt{K_r(f)} & \text{for } K_r(f) > s_a \\ 0 & \text{Otherwise} \end{cases}$$

(2)

$$y(t) = \mathcal{F}^{-1}\{\hat{W}(f)X(f)\}$$

(3)

An example that illustrates the utility of the spectral kurtosis filtering method is presented in Figure 4. The example vibration signal was acquired from a wind turbine drivetrain tested on a dynamometer test cell [2]. The raw vibration waveform plotted in the top plot in Figure 4 (b) does not indicate any impact or fault signature and the signal is quite broadband. However, using the Wiener filter based on the spectral kurtosis method, a filter plotted in Figure 4 (a) is recommended for finding the impulsive signature hidden in the vibration signal. The filtered signal is shown in the bottom plot in Figure 4 (b) and one can observe a significant amount of impacts that are occurring in the filtered vibration signal. The kurtosis value for the raw signal was only 3.39 while the filtered signal had a kurtosis value of 169. Further examination of the
signal showed that the vibration impacts were periodic with respect to the planetary gearbox carrier rotational speed. It was documented in the failure report that both the ring and sun gears in the planetary gearbox were damaged [2]. This example highlights the importance of applying more advanced signal processing and filtering methods for more accurate detection of mechanical component degradation.

![AN4 Wiener Filter](image1)

![AN4 Time Plot](image2)

![AN4 Filtered Time Plot](image3)

Figure 4: (a) Wiener Filter Based on Spectral Kurtosis, (b) Raw and Filtered Signal

2.4.2 Frequency Domain Analysis

One of the more established signal processing methods is the use of the Fourier Transform for analyzing the measured waveform in the frequency domain. In particular, for vibration signals, there is a wealth of information one can observe and extract from the vibration spectrum. For vibrations related to shaft and gear signals, in which the vibration is periodic with respect to the shaft rotation, the use of FFT analysis is an effective tool. An example of how the FFT can show the spectrum content related to worn teeth on a gear wheel is illustrated in Figure
5. In this example, the vibration spectrum for the gearbox in the baseline condition is shown in the top plot, while the bottom plot shows the spectrum for the degraded gearbox.

![Figure 5: Vibration Spectrum - Top Plot – Baseline Spectrum, Bottom Plot - Degraded Gearbox](image)

For the baseline condition, one can observe clear peaks at the gear mesh frequency and harmonics. Regardless of the gearbox condition, due to imperfect manufacturing of the gears, there will normally be spectrum content at the gear mesh frequency and harmonics. However, the sidebands around the gear mesh frequency are more indicative of the gear health condition. The bottom plot indicates that the gearbox with the degraded condition has sidebands spaced at 30Hz around the gear mesh frequency. The magnitude of the sidebands is larger than the gear mesh frequency peak for the degraded gearbox. This is a clear indication of worn gears, and the spacing of the sidebands provides an indication of which gear in the drivetrain is having problems. In this instance, the worn gear is on the high speed shaft. The high speed shaft has a frequency of 30Hz which is equal to the sideband spacing. This example interpretation of the
spectrum highlights the type of information that one can extract from frequency domain analysis. Filtering and demodulation methods can also be applied prior to calculating the spectrum for more robust and incipient detection of mechanical component degradation.

2.4.3 Time-Frequency Based Methods

The disadvantage of looking only at the frequency domain representation of the measured signal is that the temporal information is not considered. For non-stationary signals, in which the magnitude or frequency change over a short period, the use of a time-frequency distribution to represent the signal energy with respect to time and frequency is a more appropriate method. A review of the various time-frequency methods including the short time Fourier transform (STFT), the continuous wavelet transform (CWT), the Wigner Ville distribution, and the Hilbert Huang transform, is beyond the scope of this study. However, this brief review highlights an example time-frequency distribution and also provides some example statistical processing of the signal that can be used for feature extraction.

![Figure 6: Spectrogram of Example Signal with Step Change in Frequency](image)
For illustrating a time frequency signal based on the STFT, a simulated signal was considered, which consisted of a sinusoidal signal at 10Hz that changed in frequency to 200Hz halfway through the time period. The STFT spectrogram of the signal, $S(t, \omega)$, is provided in Figure 6, in which one can clearly observe the change in frequency that occurs at 5 seconds. As one can observe, it is difficult to have precise identification of the exact time instance when the frequency changes, since there is a tradeoff between frequency and time resolution with respect to the block size [20].

To further analyze the time-frequency representation, one can normalize the signal by the total energy in the signal to form a statistical distribution, $P(t, \omega)$, as illustrated in Equation 4.

$$P(t, \omega) = \frac{S(t, \omega)}{\sum \sum |S(t, \omega)|} \tag{4}$$

The marginal distributions, $P(\omega)$ and $P(t)$, can also be calculated if one integrates with respect to time or frequency as illustrated in Equation 5-6.

$$P(\omega) = \int P(t, \omega) dt \tag{5}$$

$$P(t) = \int P(t, \omega) d\omega \tag{6}$$

The marginal distributions can be used to calculate statistical measures, such as the average frequency for the measured signal, which can be calculated using Equation 7.

$$\bar{\omega} = \int \omega \cdot P(\omega) d\omega \tag{7}$$

From the marginal distribution, one can also calculate conditional distributions, such as what is the frequency given this particular time instance. The formula for calculating the conditional frequency distribution is provided in Equation 8.
From the conditional distribution, one can consider some rather insightful signal characteristics, such as the mean conditional frequency over time. This represents the average frequency value at different time instances, in which Equation 9 can be used to calculate this quantity. In addition, one could consider the variation in the frequency at different time instances. The conditional frequency variance represents the variation in signal frequency over time and can be calculated using Equation 10 [21-22].

\[
P(\omega | t) = \frac{P(t, \omega)}{P(t)}
\]

The mean and variation in conditional frequency with respect to time is further examined with another example non-stationary signal. In this example, a sinusoidal signal is initially at 2Hz, and then changes in frequency to 10Hz halfway through the time period. The conditional mean frequency for this signal is provided in Figure 7, and one can notice that this function of time matches the true frequency quite well. The conditional mean frequency has a value of 6Hz at 5 seconds, which is when the signal changes frequency and represents the average frequency between 2Hz and 10Hz. The conditional variation in frequency with respect to time is shown in Figure 8 and provides another way of examining the signal characteristics. The variation in signal frequency is quite low for the constant frequency time instances, but there is a significant large variation in frequency as one approaches the time instance when the signal changes in frequency. This implies that the variation in signal frequency could be quite useful for locating the time instance when a signal changes in frequency. Effectively, one would need to calculate scalar quantities and features from these conditional frequency functions,
considering that most of the health assessment and health diagnosis methods require a feature vector as an input to determine the system condition.

Figure 7: Conditional Mean Frequency

Figure 8: Conditional Frequency Variation Function
2.4.4 Rolling Element Bearing Feature Extraction Algorithms

A previous study on rolling element bearing health assessment and remaining useful life prediction evaluated a wealth of feature extraction, health assessment and prognostic algorithms [23-24]. After conducting the bearing failure prediction study [24], it was possible to determine the advantages and disadvantages of the different feature extraction methods for rolling element bearing condition monitoring. A tabular summary of each feature extraction method with their relative merits and disadvantages are presented in Table 3. It should be noted that this table does not include modifications or enhancements for each method; in particular, finding the optimal frequency band for the bearing envelope analysis method is an ongoing area of research [18].

Although each method has their strengths, it is from the author’s experience that the bearing envelope analysis method is the most effective and established signal processing and feature extraction method for rolling element bearing monitoring. Extracting bearing fault peak information using the traditional FFT is only effective if the bearing damage or spall size is quite severe. The use of time domain statistics provides some insight on the overall health state of the monitored system but would not provide any detailed information to help diagnosis the particular component or failure mode.

Features based on wavelet decomposition can quite effective for detecting incipient degradation, but it is difficult to attribute the change in feature magnitude to only bearing degradation. In a similar manner, features based on the empirical mode decomposition (EMD) are more suited for overall condition assessment and less suitable for only detecting bearing degradation due to the course frequency resolution. In addition, the amount of computational effort needed for the empirical mode decomposition makes it impractical for the calculation to be implemented in an embedded monitoring system.
Table 3: Summary of Feature Extraction Methods for Bearing Condition Monitoring

<table>
<thead>
<tr>
<th>Signal Processing Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Time Domain Statistical</td>
<td>Provides an overall indication of mechanical system health.</td>
<td>Limited root cause information [25].</td>
</tr>
<tr>
<td>Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. FFT Bearing Fault</td>
<td>Good for detecting bearing damage at late stages.</td>
<td>Not suitable for detecting incipient damage [26].</td>
</tr>
<tr>
<td>Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Bearing Envelope</td>
<td>Good for detecting incipient damage [27].</td>
<td>Requires a high sampling rate.</td>
</tr>
<tr>
<td>Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Wavelet Decomposition</td>
<td>Suitable for non-stationary signals and provides energy in frequency bands [28].</td>
<td>Requires selection of the mother wavelet and decomposition level.</td>
</tr>
</tbody>
</table>

Considering that bearing envelope analysis is the most established feature extraction method for bearing condition monitoring, a brief review of the method is presented. A flow chart that shows the processing steps for bearing envelope analysis is provided in Figure 9. A key step in the process is the selection of the band pass filter frequency range, in that one should select a frequency range that includes a structural resonance that is being excited by the
bearing fault impact. The output of the method is the envelope spectrum, in which one can observe the bearing fault frequency peaks. Whether the bearing fault is on the inner race, outer race, rolling element, or cage, will determine which fault frequencies or sidebands one observes in the envelope spectrum. The envelope RMS value is a feature that is suited for determining the overall bearing condition, while the bearing fault frequency peaks can be used to further diagnosis the fault type and location.

![Figure 9: Bearing Envelope Analysis Flow Chart](image)

To further illustrate the bearing envelope analysis method, a vibration signal from a wind turbine drivetrain tested on a dynamometer test cell is used, in which the gearbox was in a degraded condition with multiple bearing failures [2]. The envelope spectrum from two accelerometers, one near the intermediate speed shaft and the other one near the high speed shaft are provided in Figure 10. The failure report indicated that the intermediate speed shaft
(ISS) upwind bearing and the high speed shaft (HSS) downwind bearing both had damage on the inner race. The envelope spectrum shows strong spectral peaks at the BPFI for both the intermediate shaft bearing and the high speed shaft bearing. The inner race fault frequency peak for the high speed shaft bearing can be detected from both accelerometers but the magnitude of the peak with respect to the noise floor is much higher with the accelerometer closest to the high speed shaft. This example highlights the effectiveness of the envelope analysis method, in that even for a complicated wind turbine gearbox with multiple gear, shaft, and bearing components; one can identify spectral peaks related to the damaged bearing components.

Figure 10: Envelope Spectrum (a) ISS Accelerometer - Peaks at BPFI for ISS Upwind Bearing and HSS Downwind Bearing, (b) HSS Accelerometer - BPFI Peak for HSS Downwind Bearing

2.4.5 Gear Feature Extraction Algorithms

As for rolling element bearings, there is a diverse set of condition monitoring and feature extraction algorithms that have been developed over the years for gear components. The most popular and established gear feature extraction methods are based on time synchronous averaging (TSA) and further processing of the TSA signal. Time synchronous average is a method that uses the tachometer signal to angular re-sample the vibration signal and ensemble
average the vibration signal over a time period of one shaft revolution. Effectively, vibration that is not synchronous with respect to the shaft rotational speed is reduced and filtered out. For time synchronous averaging, the signal to noise ratio is a function of the number of averages [30]. A tabular overview of some of the more commonly used feature extraction method for gears are provided in Table 4.

Table 4: Overview of Gear Feature Extraction Methods

<table>
<thead>
<tr>
<th>Signal Processing Technique</th>
<th>Feature Extraction Method</th>
<th>What Failure Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Synchronous Averaging and Frequency Domain Analysis.</td>
<td>Sideband level and sideband ratio features.</td>
<td>Eccentric gear, misalignment, distributed type of gear fault problems [31].</td>
</tr>
<tr>
<td>3. Synchronous Average and Amplitude Modulation Signal.</td>
<td>For eccentric gear and other distributed faults, RMS and peak to peak of amplitude modulation signal can be used; kurtosis is best if there is a local gear fault problem.</td>
<td>RMS and peak to peak features are suited for eccentric gear, kurtosis is good if there is localized fatigue damage on a gear tooth [33].</td>
</tr>
<tr>
<td>4. Synchronous Average and Phase Modulation Signal.</td>
<td>Kurtosis of phase modulation signal and kurtosis of derivative of phase modulation signal.</td>
<td>Can provide early indication of fatigue damage on gear tooth [34].</td>
</tr>
</tbody>
</table>
Considering that there is a multitude of different gear failure modes that can be experienced in the field, each feature extraction method is more sensitive and tuned to a particular failure mode. For gears with an eccentric shaft, misalignment, or distributed wear, frequency domain analysis and sideband level and sideband ratio features are an effective method for detecting this type of problem [31]. However, for pitting damage on an individual gear teeth, removing the gear mesh harmonics and using the gear residual signal and various statistics such as RMS and kurtosis is the most appropriate method. Performing a narrowband filtering around a gear mesh frequency and calculating the amplitude modulation signal can be used for both an eccentric gear fault and a crack or fatigue damage on the gear tooth. The amplitude modulation peak to peak value is an effective feature for the eccentric gear fault; while the impacts and kurtosis value of the amplitude modulation signal are more indicative of a crack on the root of the gear tooth. Lastly, the narrowband phase modulation signal can provide more incipient detection of a gear tooth fatigue crack; the kurtosis of the phase modulation signal is one feature that is quite suited for the early detection of this problem.

The interpretation of the gear signal processing and feature extraction methods are best illustrated with an example; the subsequent results are from a wind turbine gearbox [2] with multiple damaged gear teeth. For the time synchronous average (TSA) frequency domain analysis, it is important to quantify the magnitude of the sidebands around a given gear mesh frequency. In order to quantify the magnitude of the sidebands, the sideband level was calculated using Equation 11. In this calculation, \( \text{SBL}_a \) stands for the sideband level, \( \text{Sb}_{a1} \) is the magnitude of the lower sideband and \( \text{Sb}_{a2} \) is the magnitude of the upper sideband. In addition, a sideband ratio was also calculated using Equation 12; this normalizes the sideband ratio by the gear mesh frequency peak. Prior work has shown this sideband ratio feature to be an effective metric for quantifying gear health since it is less dependent on the torque load [35].

\[
\text{SBL}_a = \text{Sb}_{a1} + \text{Sb}_{a2} \tag{11}
\]
Example results using the frequency domain gear features are provided in Figure 11. In this example, the sideband ratio for the intermediate speed pinion and the high speed pinion are larger in magnitude for each sample from the degraded gearbox compared to the gearbox from the baseline condition. However, there is more separation in this condition monitoring feature for the high speed pinion than what is observed for the intermediate speed pinion. The sideband ratio for the high speed gear is very similar to the baseline level and would imply that this particular gear wheel is normal using this frequency domain feature. It should be noted in
the failure report [36], that high speed pinion was observed to have severe scuffing. In addition, the failure study [36] reported that the intermediate speed pinion had severe fretting corrosion and scuffing as well. From the frequency domain method, it appears that there is a strong indication that there is damage on the high speed pinion. There also is an indication but with lower confidence of damage on the intermediate speed pinion.

An alternative method for assessing the gear wear based on the sideband level can be conducted using cepstrum analysis. The following example using cepstrum analysis is from the same wind turbine gearbox that had a damaged high speed shaft pinion [2]. The real cepstrum provides a processing method that is ideally suited for analyzing a family of harmonics in a more consolidated way than the frequency domain representation. For calculating the real cepstrum, the inverse Fourier transform is applied to the logarithm of the power spectrum as shown in Equation 13, where \( C_{xx}(t) \) is the real cepstrum and \( A(f) \) is the frequency spectrum [13].

\[
C_{xx}(\tau) = \mathcal{F}^{-1}[2\ln(A(f))] \tag{13}
\]

For mechanical systems and gears in particular, the cepstrum provides a convenient way of analyzing a series of sidebands that are spaced at a given shaft speed; comparing the cepstrum from a baseline and current state can be used to infer the health condition of each gear wheel. The example cepstrum result in Figure 12 further illustrates this aspect in which the cepstrum from the baseline gearbox is compared to the degraded gearbox. In both instances one can observe a peak in the cepstrum at 0.133s, which corresponds to 7.5Hz and the intermediate speed shaft. This implies that a family of harmonics spaced at 7.5Hz was always present in this gearbox. However, an additional peak at 0.0325s, which corresponds to 30Hz and the high speed shaft can be seen in the cepstrum of the degraded gearbox. This additional set of harmonics spaced at 30Hz for the degraded gearbox provides evidence that the gear wheel on
the high speed shaft (high speed pinion) is degraded and is responsible for this noticeable change in the cepstrum.

![Real Cepstrum of AN7 HSS Radial (Before)](image)

![Real Cepstrum of AN7 HSS Radial (After)](image)

Figure 12: Real Cepstrum, Top Plot - AN7 Baseline, Bottom Plot - AN7 Degraded Gearbox

Additional vibration features were extracted from the cepstrum at the corresponding shafts using data from both the baseline gearbox and the degraded gearbox. Example results from the cepstrum features are provided in Figure 13, in which several peaks in the cepstrum are larger in magnitude when comparing the degraded gearbox to the baseline gearbox. The cepstrum peak related to the high speed pinion is clearly larger in magnitude for the degraded gearbox. This provides an additional set of evidence that the high speed pinion is damaged.
In addition to cepstrum analysis, the narrowband amplitude and phase modulation analysis was also conducted using the vibration data from the same wind turbine gearbox that had damaged on the high speed shaft pinion. For detecting local defects such as a fatigue crack in a gear wheel, the prior work done by McFadden [34] suggested the analysis of the amplitude and phase modulation signals of the gear vibration. For performing this analysis, the synchronous average signal for each shaft is extracted. A band pass filter around a dominant gear mesh frequency is used and typically includes a number of sidebands around the gear mesh frequency peak. The Hilbert transform is then performed on the filtered signal; the modulus and phase of the analytical signal provide the envelope and phase modulation signals respectively. The amplitude and phase modulation signals were calculated for each gear wheel in this study. In addition, the kurtosis of the amplitude modulation signal and the kurtosis of the derivative of the phase modulation signal were also calculated in order to quantify the health
condition of each gear. For the parallel shaft gears, the band pass filter included 4 sidebands, while the band pass filter for the ring gear included 6 sidebands. The accelerometer on the planetary gear housing was also effectively down sampled to 200Hz prior to extracting the synchronous average for the ring gear. Sample results for this method are provided in Figure 14, in which the amplitude and phase modulation signals are plotted for the high speed pinion. As one can observe, there is significant jumps in the phase modulation signal for this gear and a high kurtosis value at 12.8. The phase modulation signal also provides additional evidence that the high speed pinion is in a degraded state.

![Image of graphs showing time synchronous average, amplitude modulation, and phase modulation signals for high speed pinion.](image)

Figure 14: High Speed Shaft Pinion AM and PM Signal, Top Plot - Time Synchronous Average, Middle Plot - Amplitude Modulation Signal, Bottom Plot - Phase Modulation Signal

2.5 Feature Selection Methods

In the feature extraction step, all the potentially relevant features are extracted from the measured signals. However, it is not prudent to include all the extracted features in the health or
prediction models, and selecting a subset of those features can provide a more robust and computational efficient model [37]. A summary of the feature selection methods are listed in Table 5, in which the consideration of which one to use, is based on the data characteristics and whether data is available from multiple health states.

Table 5: Overview of Feature Selection Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Expert Selection</td>
<td>Does not require data sets for ranking or selecting features, suitable for more established monitoring applications such as rotating machinery.</td>
<td>Requires significant experience for each application, for new applications it is very difficult to select the features based on expert knowledge.</td>
</tr>
<tr>
<td>2 Filter Methods</td>
<td>Easy to implement and allows one to have a metric to rank each potential feature.</td>
<td>Requires data sets from a healthy and degraded system, also many of the filter methods are more designed for classification applications.</td>
</tr>
<tr>
<td>3 Trendability Metrics</td>
<td>Provides metrics for determining how monotonic and consistent the failure patterns are for features.</td>
<td>Ideal for run-to-failure test and prognostic model development, however the metrics are less developed than other feature selection filter methods.</td>
</tr>
<tr>
<td>4 Wrapper Methods</td>
<td>Finds the best feature subset by including the regression or classification performance in the search routine.</td>
<td>Requires training data sets and also can be influenced by the selected classification or regression algorithm.</td>
</tr>
</tbody>
</table>
Expert selection of which features to include is particular appropriate if only data from a baseline system is available and one has sufficient domain knowledge of which parameters are physical related to the system or component degradation [38]. However, selecting the features for new applications in which there are less established methods or an understanding of the physical degradation process makes it very difficult to selecting the features based on expert knowledge.

Filter based methods offer the ability to rank the features based on their ability to discriminate between the various healthy and degraded health states; fisher criterion, information gain, correlation coefficient, and other metrics are available for performing the feature ranking [39]. The caveat to using filter based methods is that one must have sufficient data from the system in multiple health conditions, with the minimum requirement of having data from both a healthy state and a degraded state. For prognostic applications, the use of conventional feature selection metrics might not be appropriate since they are more designed for measuring the separating between clusters. Metrics that rank a set of features based on whether they are monotonic, have a repeatable degradation pattern, and a consistent failure value are more suited for determining the best feature set for a prediction model [5]. The previously described feature selection methods only provide a ranking of the individual features; however they do not consider the interaction between the features nor do they determine the number of features to include in the model. Wrapped feature selection methods iteratively select the optimal subset based on the performance of a regression or classification model [40]. Wrapper methods require more computation and can be influenced by which regression or classification algorithm is used for determining the feature subset to include in the model.

2.6 Health Assessment Methods

There are several algorithms that are used to fuse multiple features into a single health metric or an anomaly detection output; a listing of the more commonly used algorithms for
assessing the machine health are provided in Table 6. Perhaps the most straightforward method is to use a health metric based on a summation or weighted summation of the feature values. This health metric is simple to implement and statistical thresholds can be derived based on the distribution of the health value [41]. Various distances from normal health metrics can be used for determining the health condition of the monitored system or component. Mahalanobis distance and principal component analysis based Hotelling’s $T^2$ statistics are distance metrics that incorporate the covariance relationship among the variables; however Euclidean and other distance metrics are also commonly used [42]. Although the distance based health values are very applicable if only baseline data is available, the distance method only accounts for a change in normal and not the direction. This has potential drawbacks, in that a system can have lower than normal vibration or temperature and this would still trigger a higher health value and an anomalous condition.

A simple but effective approach for anomaly detection is the use of statistical hypothesis testing. Sequential probability ratio test, rank permutation test, and T-test, are an example of some of the more commonly used hypothesis test for anomaly detection [43]. Other anomaly detection based methods include a one-class classifier, such as the support vector data description algorithm. In this algorithm, data from the baseline condition is used for form a hypersphere, in which new data outside this boundary is considered an anomaly [44]. One particular disadvantage of the support vector data description algorithm is the selection of the kernel function, in which there is not substantial guidance in the literature on which kernel function to use for a given application. Regression or neural network based methods are particular effective if sufficient data is available for developing the regression models. The neural network or regression model can be used to provide a mapping between the feature values and a health value or defect size [45]. However, the ability for the model to generalize usually requires training data sets from multiple monitored systems.
<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighed Summation of Features</td>
<td>Simple to implement, also easier for setting thresholds.</td>
<td>Does not learn from any data sets from a degraded or failed component or system.</td>
</tr>
<tr>
<td>Distance from Normal</td>
<td>Requires only baseline data sets for training the algorithm.</td>
<td>Does not account for whether the features are lower or higher than expected.</td>
</tr>
<tr>
<td>Statistical Hypothesis Testing</td>
<td>Simple to implement and computationally efficient.</td>
<td>Data might not fit assumed distribution for the hypothesis testing.</td>
</tr>
<tr>
<td>Regression Methods</td>
<td>Provides a mapping between the feature values and the output health value.</td>
<td>Requires an output value that is related to the health condition of the system.</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>Can be used as a non-linear regression model for mapping the feature values to the health condition.</td>
<td>More difficult to train and troubleshoot than regression based methods.</td>
</tr>
<tr>
<td>One-Class Classifiers</td>
<td>Support vector data description algorithms can provide a boundary for anomaly detection.</td>
<td>Requires experience on selecting the appropriate kernel function.</td>
</tr>
</tbody>
</table>
2.7 Health Diagnosis Methods

For root-cause analysis and diagnosis, there are many different methods and algorithms for performing this task; a sample of some of the more commonly used methods is listed in Table 7. Incorporating engineering knowledge and experience into the diagnostic algorithm makes the use of fuzzy membership functions and rules an attractive technique [46]. However, it becomes more challenging to use fuzzy based diagnostic algorithm for new applications in which there is not sufficient experience on the failure modes and their signatures.

<table>
<thead>
<tr>
<th></th>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fuzzy Membership Rules</td>
<td>Can include engineering knowledge and experience in the diagnostic algorithm, more flexible to adjust the algorithm compared to machine learning methods.</td>
<td>Requires experience for determining the rules and membership functions, rules might have to be adjusted to account for variation among the monitored units in a fleet.</td>
</tr>
<tr>
<td>2</td>
<td>Machine Learning Classifier Algorithm</td>
<td>Classification algorithms can learn the relationship between the features and the output classification label.</td>
<td>Requires data from each fault class for training the algorithm, ability for the algorithm to generalize is not straightforward.</td>
</tr>
<tr>
<td>3</td>
<td>Bayesian Belief Network</td>
<td>Can model the cause and effect relationship between the feature values and the system health states.</td>
<td>Determining the BBN structure requires experience or the use of less established methods for learning the network structure.</td>
</tr>
</tbody>
</table>
The use of a classification algorithm is a popular alternative if there is data from multiple health states including a baseline condition and several of the different failure modes that can occur. The use of neural networks, support vector machines, and Naïve Bayes algorithm are some of the more common classification algorithms used for machine condition monitoring [47]. By learning the relationship between the extracted features and the baseline and failure signatures, the classification method can accurately diagnosis and label the health condition from the monitored system.

Other methods for diagnostics include the use of a Bayesian Belief Network (BBN) which provides a network representing the casual relationship between the measured variable and the different failure modes or system conditions that can occur [48]. In order to apply a BBN network for machine diagnostics, the feature values have to be changed from continuous values to discrete states. In addition, determining the BBN structure for a given application requires sufficient engineering domain knowledge. Learning the BBN structure from substantial training data is one option, but would still require someone with domain knowledge to determine whether the network is logical and makes physical sense.

2.8 Remaining Useful Life Prediction Methods

A sample of the more commonly used remaining useful life prediction algorithms are presented in Table 8, along with the advantages and disadvantages of each method. Curve fitting based methods are a relatively simple method to apply, in that they do not require a substantial amount of training data or a detailed physical model that describes the fault progression. However, the curve fitting based prediction results require an appropriate selection of the curve fitting model, which requires experience, historical data, or some insight on the failure mechanism [24]. Neural network or regression based methods can directly map the feature values and provide an estimation of the remaining useful life [49]. These methods
require substantial training data for learning this relationship and obtaining multiple run-to-failure data sets is not feasible in many applications.

Table 8: Overview of Prognostic Algorithms

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Curve Fitting Methods</td>
<td>Simple to implement, does not require substantial training data sets.</td>
<td>Results are dependent on selecting an appropriate curve fitting model form, does not directly incorporate a physical degradation model.</td>
</tr>
<tr>
<td>2 Neural Network or Regression Methods</td>
<td>Can learn the relationship and provide a mapping between the feature pattern and the remaining useful life.</td>
<td>Requires several run-to-failure data sets for learning this relationship, also does not incorporate the physics of the degradation process.</td>
</tr>
<tr>
<td>3 Stochastic Filtering Methods</td>
<td>Incorporates the failure physics and can handle uncertainties in the modeling and sensor data.</td>
<td>Requires a physical model to describe the failure mechanism. Also in main instances the defect level is not directly measurable.</td>
</tr>
<tr>
<td>4 Similarity Based Prediction Method</td>
<td>Accurate and can account for different degradation patterns or initial degradation conditions.</td>
<td>Requires several run-to-failure data sets for developing a degradation pattern library.</td>
</tr>
</tbody>
</table>
A similarity based prognostic algorithm is a unique method that matches the previous degradation patterns to the current degradation pattern of the monitored system [50]. The similarity based prognostic algorithm can be quite accurate. However, it requires several run to failure data sets in order to obtain a library for performing the degradation trajectory matching. As a contrast to the previously described data-driven prognostic algorithms, the incorporation of a physical model of the failure mechanism with a stochastic filtering algorithm is an effective approach. Whether the fault propagation dynamic equations are linear or non-linear and whether the measurement noise is Gaussian, can be use to select whether a Kalman filter, or extended Kalman Filter or particle filter are used [51]. Applying model based prediction algorithms using stochastic filtering does have some potential challenges. Only for a subset of applications does one have established models for describing the failure mechanism.
CHAPTER 3: PRIOR METHODS USED FOR THE PRONGOSTICS AND HEALTH ASSESSMENT OF A MULTI-REGIME SYSTEM

3.1 Introduction

Monitoring the health state of a system operating under multiple regimes has been a relatively recent research topic in the prognostics and health management community, considering that 65% of the 34 papers surveyed were published within the last 4 years. The challenge of monitoring a system under changing operating conditions has led to a multitude of different approaches which are highlighted in Table 9. The approaches can be divided into four different categories, including local models, normalization methods, statistical distribution approaches, and remaining useful life (RUL) estimation algorithms. Of the four categories, RUL prediction algorithms for multi-regime system requires the most additional research, in that these methods have only been demonstrated in a test-bed or simulation setting due to limitations in both the data-driven and physics of failure models. The data-driven methods require an abundant amount of run-to-failure training data which are difficult to obtain in practice. The physics of failure models would work well for systems that are operating in a single operating condition. However, for multi-regime systems they require an estimate of the future usage which is difficult to accurately predict. Thus, the focus of the literature survey is on health assessment for multi-regime systems; this is also aligned with the research contribution of this dissertation.
Table 9: Summary of Literature on Multi-regime Health Monitoring and Prognostics

<table>
<thead>
<tr>
<th>Approach</th>
<th>Algorithm / Method</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Models</td>
<td>PCA [52], distance metric [53], regime segmentation [54], fuzzy membership [55],</td>
<td>Diesel engine [52], gearbox [53,56], gas turbine</td>
</tr>
<tr>
<td></td>
<td>residuals in regime bins [57]</td>
<td>[5], wind turbine [57-58]</td>
</tr>
<tr>
<td>Normalization</td>
<td>Physical equations [59-60], linear regression [61-62], neural network regression [63-65],</td>
<td>Sensor faults [59,71], gas turbine [60,67,75],</td>
</tr>
<tr>
<td></td>
<td>auto-associative models [66-72], first principal physical models [73-76]</td>
<td>gearbox [61,66], roller coaster [62], wind turbine [63], process monitoring</td>
</tr>
<tr>
<td>Distribution Methods</td>
<td>Mixture of Gaussians overlap method [77-78], distance from cluster calculation [79-80]</td>
<td>Wind turbine [78], chiller compressor [77], process monitoring [79-80]</td>
</tr>
<tr>
<td>Remaining Life Estimation</td>
<td>Data driven similarity method [50,81], Stochastic filtering physics of failure model [83-34]</td>
<td>Gas turbine [50,81], bearing [82], milling [82], battery [83], gearbox [84]</td>
</tr>
</tbody>
</table>

The literature review summary in Table 9 highlights the diverse set of algorithms that are used in the local models, normalization approaches, distribution methods and RUL estimation methods. Techniques ranging from distance metrics, principal component analysis monitoring, auto-associative models, first principle physical models, Gaussian mixture models, and similarity based prediction methods have been used to solve these multi-regime problems. The
applications are equally diverse, ranging from diesel engines, mechanical gearboxes, sensor health estimation, roller coaster guide wheels, wind turbines, batteries, and aircraft engines. Although those applications have their own unique challenges, they also represent a common scenario in which the condition monitoring system is further tasked with estimating the component or system health under changing operating regimes.

3.2 Local Model Approach

One approach to the multi-regime problem is to divide the operating space into smaller sub-regimes and have a health model in each regime. A graphical representation of this concept is shown in Figure 15, where in this example there are two operating variables. The selection of how to divide the operating space is application dependent and can based on ones engineering experience or knowledge about the system. A more in depth discussion on the different type of partitions for the local model approach are provided by Wang et al. [54]. This type of local modeling approach has been demonstrated in various applications. A previous study by the author [52] used local auto-associative neural network and principal component analysis models for monitoring the health condition of a diesel engine. There were no false alarms but there was a fault detection rate of 83% for the highest engine operating speed compared to a 73% detection fault detection rate for the lowest engine operating speed regime. A distance from normal metric in each regime was used by Atat et al. [53] for monitoring the health of bearing, shaft, and gear components in a gearbox. The method used in [53] resulted in the most accurate health monitoring results for the student division of the PHM 2009 Data Challenge; however it was noted that the vibration fault signature in certain operating regimes was much more pronounced.
An extension of the local modeling approach was presented in [55], in which fuzzy membership functions were used to handle transitions between the different operating regimes. The membership functions along with the auto-associative neural network models provided a weighted residual calculation for the aircraft engine sensor variables. The method showed promise, however the initial evaluation was on simulated data from a high fidelity aircraft engine model. Additional work related to gearbox condition monitoring was conducted by Klein et al. [56], which used local models and fused the results from each regime and from multiple sensors to make the final diagnosis for a given gear or bearing component.

This type of local modeling approach has also been quiet popular for wind turbine condition monitoring. Wind turbine power curve residuals in each wind speed bin were used by Uluyol et al. [57]; however, the selection of the bins is not trivial and requires engineering experience for the method to work effectively. The local model approach was also considered in [58] for a vibration based monitoring method for a wind turbine gearbox. The monitoring results in [58] noted that the fisher discriminate value was much higher in certain output power regimes.

There are two potential drawbacks to the local model approach for multi-regime health monitoring. The first aspect is that the degradation signature is more pronounced in certain
operating regimes than in others. The diesel engine health monitoring case study noted that better results were obtained in the highest operating speed [52], while the gearbox case studies [53, 58] reported better discrimination under certain loading or operating speed conditions. The local model approach does not effectively consider this regime dependent signature aspect and would require one to fuse the health values from each regime in order to infer the health state of the system. However, fusion has its own set of difficulties and one concern would be that the fusion method could dilute the results if the degradation signature is only present in a few of the operating regimes. The other potential challenge in using the local modeling approach is when there are more than three operating variables. Partitioning the operating regimes when there are several operating variables is quite difficult and makes the local model approach quite cumbersome in this situation.

3.3 Normalization Methods

3.3.1 Physical Equations

The features related to the degradation of the component or system are also affected by the operating conditions. Differentiating between machine degradation and a change in operating parameters would require some type of normalization process. Perhaps the most straightforward approach is to use known physical relationships between the operating parameters and the measured signals or features. An example of this normalization approach was used by Cassity et al. [59] for anemometer sensor health monitoring. The normalization method in [59] used the power law to relate the wind speed measurements at different heights and took into account the logarithmic wind profile law by calculating an $R^2$ value after applying a log transform to the data. This normalization approach for wind speed sensor fault detection proved effective and resulted in the third best score in the 2011 PHM Society Data Challenge.
This type of equation based normalization approach is perhaps the most common for gas
turbine and aircraft engine condition monitoring. Correcting the pressure and temperature
measurements for ambient conditions is a common practice and most existing aircraft engine
monitoring systems provide corrected sensor parameters [60]. Although this type of equation
based normalization approach is quite effective, it is only limited to applications that have these
established physical relationships for normalization the data. In addition, the equations are
usually empirical based and were thus fitted to some type of experimental or field data. In
certain instances it might be necessary to update these equations with data from the actual
component or system that the condition monitoring system is being developed for.

3.3.2 Regression

A mapping between the operating variables and the calculated degradation features can be
modeled by different types of regression approaches, with a linear regression model being an
approach choice in certain applications. The normalized features or the calculated parameters
of the linear regression model can then be used to help discriminate between the healthy and
degraded state of the monitored system. A linear regression based normalization approach was
considered by Bartelmus et al. [61] for monitoring the health state of a planetary gearbox in a
mining application. The condition monitoring feature consisted of the sum of the gear mesh
frequency (GMF) peak and its first 9 harmonics, and a linear regression model was developed
that related this feature to the rotational speed. The slope of the regression equation provided a
clear way to discriminate between a planetary gearbox in a health state and one with distributed
damage on the gear teeth. However, it was noted that a nonlinear regression model might be
considered for future work considering that the damaged gear teeth could cause an increase in
backlash [61].

Additional work in terms of regression based normalization was conducted by Sohn et al.
[62] for monitoring the health condition of a roller coaster guide wheel. Degradation of the
polymer layer of the roller coaster guide wheel could be a safety, rider comfort, or maintenance issue, and this study was aimed at monitoring the health signature of the wheel using accelerometers mounted near the track. A normalization method is necessary since the weight of the passengers could greatly affect the measured vibration response and mask the signature of a degraded guide wheel. In this study, auto-regressive models were considered for normalizing the load effect. However, the results indicated that additional work was necessary for improving the normalization method and providing a more confident detection of a degraded wheel. Although linear regression or linear predictor models offer one approach for normalization the condition monitoring features, the relationship between the operating variables and the features are typically nonlinear. Thus despite their simplicity, linear regression models are not usually the best approach for performing the normalization in these multi-regime health monitoring problems.

3.3.3 Nonlinear Regression

Nonlinear regression models, such as neural networks, can effectively provide a nonlinear mapping between the operating variables and the condition monitoring features. A comparison study between regression and neural network models was considered for a wind turbine application [63]. Linear and nonlinear regression models were developed using supervisory control and data acquisition (SCADA) wind turbine data. Residual calculations and statistical thresholds were used to detect an anomalous health state of the wind turbine. The study concluded that neural network models outperformed the linear regression models, but that the neural network models can be more difficult to interpret. A combined neural network model and principal component analysis monitoring method was considered in [64] for process monitoring. The residuals replaced the typical feature input into the standard principal component monitoring method using Hotellings’ $T^2$ statistic and the residual square prediction error statistic (SPE). Although the method was effective, it appears to be a cumbersome solution that could
effectively be replaced by using a single auto-associative neural network model. An additional study considered the use of a radial basis function neural network for sensor fault detection in a chemical reactor system [65]. The neural network models provided predicted and residual values for each sensor, and the sensor fault detection results in this study were quite promising.

Although the nonlinear regression and neural network models offer improved results compared to the linear regression models, they have some potential drawbacks. In many instances, one wants to calculate predicted values and residuals for each feature or sensor variable. However, the neural network model would typically be configured to have multiple inputs but only one output sensor variable. Auto-associative models would be a more elegant solution, in that they effectively account for the nonlinear correlation relationship among the sensor variables and provide a predicted value for each sensor input.

3.3.4 Auto-Associative Models

Auto-associative models is one of the more popular methods for modeling the correlation relationship between the sensor and feature data and normalization the effect of the operating variables. This type of auto-associative modeling approach was combined with synchronous averaging and statistical vibration features for monitoring the health condition of the individual gears in a helicopter gearbox [66]. The results were quite promising but the study was limited to seeded fault testing on a test-rig and did not consider whether the signature would be more pronounced under different speed or loading conditions. Auto-associative neural network models were also applied to an industrial gas turbine application for calculating residual values for each of the sensor variables [67]. In this gas turbine case study, a wavelet de-noising method was used as a pre-processing step to provide less noisy inputs for training the auto-associative neural network model. A review of various normalization methods was provided by Sohn [68]. The review summarized some of the advantages of using an auto-associative model to learn the nonlinear correlation present in the signal features, and applied the method to
monitor the health condition of civil structures. Nonlinear auto-associative model are also quite popular for monitoring complicated chemical processes, in which some of the earlier work was done by Thissen et al. [69]. The use of a kernel principal analysis (KPCA) method was used for process monitoring; the KPCA method showed improved results compared to the PCA method for monitoring the waste water treatment process [72]. However, kernel principal component analysis is difficult to interpret since the variables are mapped into a larger feature space and are less intuitive than the auto-associative predicted sensor values and residuals. Additional case studies that use an auto-associative model include diesel engines [70] and a power generation sensor fault detection application [71]. The diesel engine study [70] showed that the nonlinear auto-associative model performed better than the linear PCA model. This can be expected considering that one might expect a nonlinear correlation relationship in a system as complex as a diesel engine. The sensor fault detection study [71] also had promising results, but suggested that it might be necessary to retrain the model over time or if maintenance is performed.

3.3.5 First Principal Models

Using first principals to model the dynamics of a system provides an alternative physics based approach to account for the effect of the operating variables. Ideally, one would prefer to have an accurate model of the physics and dynamics for each monitored system or component; however such models are not always available or are very costly to develop. An example of this model based approach was demonstrated for an electro-mechanical actuator [74]. In this study, a Simulink model of the system was constructed for calculating residuals and also for running different parameter estimation methods. Estimation of the friction damping coefficient, the gear stiffness, and torque constant were used to determine the health condition for the simulated bearing seizure fault on the actuator system. Although the results were promising, the
demonstration was limited to an experimental test-bed. In addition, one concern in the study was whether the parameter estimation method was tuned to only a few of the failure modes.

A first principles approach was also considered for an aircraft engine application [75]. The residuals calculated by the physical model performed better than the data-driven models that used an auto-associative neural network or a PCA based approach. However, the auto-associative neural network model provided only slightly worse than the physical model in this application. A physical model and a particle filter for fault detection was also used by Daigle et al. [76] for a spacecraft application. The particle filter method showed good results but the study was limited to only simulation data. If one has an accurate model of the dynamics of the system, then a physical model appears to be one of the best approaches for monitoring multi-regime systems and accounting for the effect of the operating variables. However, auto-associative models can offer slightly worse performance but can be applied to applications in which an accurate physical model is not available.

3.4 Distribution Methods

Another approach to the multi-regime health estimation problem is to have a statistical model of the feature distribution. Health values can be calculated based on the overlap or distance value between the current feature distribution with respect to the baseline feature distribution. Such an approach was considered by Liao et al. [77] for monitoring the health of a chiller compressor. The proposed methodology included running a fixed cycle feature test under different load settings and extracting features during the transient portions of the test. A Gaussian mixture model (GMM) was used to model the feature distributions from a baseline condition and an overlap calculation was used to assess the current health condition of the chiller. Although the GMM method provided a way to discriminate between a healthy and degraded chiller compressor, it was noted that the signature was more distinct at 25% of the working load when compared to 100% of the working load. Additional work using the GMM
overlap method was performed by Lapira et al. [78] for wind turbine performance monitoring. The study compared the degradation monitoring results using GMM, a neural network residual method, and a self-organizing map distance metric. The GMM method provided the most consistent degradation trend compared to the other two methods for this wind turbine application. A mixture model approach was also used for process monitoring in [79] and a batch process in [80], in which a distance calculation from a cluster centroid was used for detecting anomalies. Although the statistical approaches such as GMM have merit, one potential disadvantage of this method is that it does not effectively account for the aspect that the signature could only be present in a subset of the operation conditions.

3.5 Prognostic Methods for Multi-regime System

3.5.1 Data Driven Prognostics

Although estimating the health state of a multi-regime system is already challenging, some recent research methods have demonstrated the ability to accurately predict the remaining useful life of monitored systems. The use of a trajectory similarity based prediction method was first proposed by Wang et al. [50] for an aircraft engine application and resulted in the top score for the 2008 PHM Data Challenge Competition. By using the prior run-to-failure degradation trajectories and accounting for the time lag, the method could accurately estimate the remaining useful life of an aircraft engine operated in six discrete operating regimes. Although the algorithm had very good results, the difficulty in applying this method in practice is the amount of training patterns needed. An extension of the similarity based method was proposed by Hu et al. [81], in which an ensemble of data driven algorithms were fused together for providing the final predicted life output. The ensemble based method consisted of five different algorithms; in which three of the algorithms were modified versions of the similarity based prediction method. The results from this study showed that a weighted output of the algorithms provided superior results than any single prediction algorithm. The ensemble based prediction method was
evaluated with a simulated aircraft engine run-to-failure data set and a simulated run-to-failure
data set for a set of power transformers. Whether the ensemble method would generalize to a
fielded system would require more validation.

Although the similarity based prediction method outperforms many other remaining useful
life estimation methods, one concern with this method is the number of training data sets
required for the algorithm to perform well. The author evaluated the similarity based prediction
method using a publically available simulated aircraft engine run-to-failure data set [85]. The
data set consisted of 6 discrete operating regimes, and 260 run-to-failure training data sets and
259 test data sets. The data set is quite similar to the 2008 PHM Data Challenge Competition.
The only difference is that for this public data set, the engine only has one failure mode (high
speed compressor efficiency). The prediction accuracy of the similarity based prediction
method as a function of the number of training data sets is provided in Figure 16 using the mean
absolute error and the PHM 2008 Data Challenge scoring metric. As a basis for comparison,
the highest score from the PHM 2008 Data Challenge Competition was a score of 12.96 per
unit, although it was from a slightly different data set [50]. One can observe that the
performance of the similarity based approach is directly linked to the number of training units.
The prediction results are undesirable if too few of training units are used since this diminishes
the ability of the method to effectively perform the degradation pattern matching. It does appear
that reasonably accurate results can be obtained if at least 40 degradation patterns are used.
The results are at a comparable level to the PHM 2008 Data Challenge top ranking score if 100
or more degradation patterns are available for the similarity based prediction method.
3.5.2 Model Based Prognostics

The amount of training data required for an accurate data driven prognostic model restricts its application for certain problems. A physics of failure model that describes how the degradation or fault grows over time represents another approach for estimating the remaining useful life of a system or component. However, there is a tradeoff in terms of accuracy and the development time. Physical models are more appropriate when highly accurate models are needed to minimize uncertainty. An example of this type of application is an unmanned air vehicle (UAV) in which the estimation and prediction of the battery state of charge and health is critical for achieving the different flight objectives and missions. A particle filtering framework for estimation and prediction of battery state of health was demonstrated by Saha et al. [83] for a
UAV application. The method took into account uncertainties in the future flight and load profile of the battery in order to provide a point estimation and distribution associated with the battery state of health. A similar particle filtering approach was also considered for estimating the spall size and damage propagation of a critical mechanical component for a rotorcraft application [84]. The proposed framework [84] had to consider the future loading profile of the mission in the remaining useful life estimation. Despite this initial research on physics based remaining useful life estimations, the research work was demonstrated on a smaller scale system or test-bed. Based on the surveyed result and the limitations of both data driven and physics based prediction methods, there is still a substantial amount of research needed in order to achieve accurate remaining useful life estimation for multi-regime systems.
4.1 Problem Statement and Assumptions

Consider a system operating under a set of operating conditions, in which the operating conditions are defined as a set of n variables denoted by \( (o_i | i=1,\ldots,n) \). Consider also a set of k condition monitoring signals, denoted by \( (s_i | i=1,\ldots,k) \), in which the set of condition monitoring signals could be from add-on sensors or existing controller signals available on the system. Typical condition monitoring signals include vibration, temperature, motor current, and pressure. A set of features are extracted from the measured condition monitoring signals, in which m features are extracted and are given by \( (f_i | i=1,\ldots,m) \). The operating variables could be ambient conditions, operating speed, load, or any potential variable that influences the condition monitoring features. However, the operating variables do not offer insight on the health condition of the monitored system and can only be used to infer the operating state of the system.

The task is to infer the health condition denoted by a health value \( h_i \), of a complex system operating under a set of operating conditions, using the condition monitoring features and the measured operating variables. The health index can be used to discriminate between a healthy and undesirable system condition, in which a corresponding threshold \( t_{hi} \), is used with the corresponding health value. In addition, the health value can be monitored over time and used to trigger other processing modules, such as a diagnosis or prediction algorithm. The proposed monitoring approach for generating a health index is based on the following assumptions:
i. The system has a fault or degradation signature that is more pronounced and observable in a subset of the operating conditions.

This implies that a subset of the operating regimes, denoted by \( (O_c \subset O_s) \), is more conducive for detecting the degradation signature. In this case, the features in this operating condition provide more separation between the normal and degraded condition. An example could be a sensor fault detection example in which the degradation signature is intermittent and not present in all the operating conditions [86]; this case study is further explored in Section 6. Additional examples include mechanical systems in which a gear defect or imbalance shaft could be easier to distinguish under higher load or speed regimes.

ii. The system operates in several of its common operating regimes during a short time period.

Consider a monitoring period denoted by \( T_i \), in which most of the operating conditions \( O_s \) of the monitored system are experience during that time frame. The monitoring period is application dependent but should represent a time frame in which the health metric could provide enough of a time window for taking the appropriate action to perform maintenance on the engineering asset. A monitoring period could represent a shift in a manufacturing example [77], a day or up to a week for a ground vehicle, and a similar time frame for wind turbines [63]. Although this would imply that an instantaneous health value is not calculated, this is necessary given that the fault signature is only present in a subset of the operating conditions.

iii. There is not sufficient prior knowledge of which operating regimes are the most conducive for detecting the degradation signature in the monitored system or component.

Ideally, if one had sufficient knowledge on which operating regime the fault signature was the most observable, then they could simply monitor the system in that operating regime. However, for complicated systems this is not known a priori. In addition, conducting a series of seeded
fault experiments in a test-bed for each operating regime would be cost prohibitive and also might not be able to generalize to the larger system that is used in the field [87]. A methodology is needed for processing data from a system operating under multiple operating regimes that autonomously selects the most conductive operating regime for monitoring the system health state.

iv. Only data from the system in a nominal healthy state is available.

It is a reasonable assumption to assume that there is sufficient data from a system in a healthy state, in that this data should be significantly easier to obtain then data in which the system is in a degraded state. This is a common scenario, in which the monitoring system has to be developed from only baseline data [88]. It is not reasonable to assume that data from a healthy state and the different failure modes would be available for training health monitoring algorithms using supervised regression or classification techniques.

v. There is some basic prior knowledge for interpreting the residual values and formulating a figure of merit health metric

This aspect is explored in more detail for the sensor health monitoring and gearbox condition monitoring case study. For the sensor health monitoring case study, there is prior knowledge that the fault signature is based on the lagging wind speed reading. Focusing on the residual values that are negative (lower than expected) can be used to exploit this prior information about the lagging wind speed sensor [86]. For monitoring mechanical systems using vibration, residuals that are positive can be used during the cluster selection and figure of merit health value calculation. For vibration based monitoring, vibration larger than normal would be indicative of a fault [89], and negative residuals (less than normal) would not imply a problem but just normal variation.
vi. There are no features that can be used to discriminate between a normal and degraded condition that are independent of the operating regimes such that no further processing is required.

Ideally, if there was a feature that was independent of the operating conditions and could also be used to discriminate between a healthy and degraded state, then there would be no need to develop an intelligent data processing algorithm. The condition monitoring features are indirect measurements that can be used to infer the system health, and in most instances, are also influenced by the operating conditions [61].

vii. The selected condition monitoring features have a nominal correlation relationship,

The assumption that the condition monitoring features have a degree of correlation is based on the rationale for using an auto-associative neural network to modeling this correlation relationship in the features. If the features are independent or have a low degree of correlation, then an auto-associative model would not be appropriate and other types of algorithms might be more appropriate; such as a first principals model for provide predicted and residual values for each signal. This aspect of the performance of the auto-associative residual processing module and the amount of correlation exhibited in the measured signals is explored in Section 4.5.

4.2 Intuition

There are several applications in which one would postulate that the fault signature is more noticeable or observable in a particular operating regime. For ground vehicle applications, one might only observe a particular engine or transmission problem at a certain engine rpm or driving speed. For aerospace applications such as an aircraft engine, the engine degradation might be more noticeable in a particular flight regime such as take-off or cruise [46]. For rotating machinery, gear, shaft, and bearing faults could be easier to detect under higher load or
higher rotational speed [90]. Previous studies regarding sensor health detection have also commented that the signature is intermittent and not present in all the operating regimes [91].

Figure 17: Concept of Degradation Being Present in a Subset of the Operating Conditions

From these examples, it appears that these applications face a common problem, in that the system is operating under multiple regimes and operating conditions, but the signature is only prevalent in a subset of those conditions. The general concept of this type of problem is illustrated in Figure 17, which illustrates that there is a regime monitoring map which would highlight which conditions are best for detecting this degradation signature. For this type of regime dependent degradation signature problem, it appears necessary to focus on a particular regime subset for calculating a health index and determining the appropriate set of maintenance actions. The processing modules necessary to develop an automated health monitoring algorithm for these types of systems with regime dependent signatures is the following:

i. A residual processing module

Calculating and comparing predicted feature or sensor values with actual measured values is one way of normalizing the affect of the operating regime variables. From ones intuition, certain features might increase under certain operating conditions (e.g.
higher load or speed); however it is important to look at how the feature values compare with the expected response under those operating conditions.

ii. A clustering algorithm

The clustering module would further process the residual values into 2 or more clusters. The hypothesis is that the signature is present in only a subset of the operating regimes and thus there is a cluster in the data that represents this operating regime subset. Since the objective is to develop a health metric that is focused on that subset, there should be criteria for selecting which cluster to focus on. Some prior knowledge about the application can be use to guide which cluster to select and use for calculating the health value. The aspect of cluster selection is discussed in more detail for the case studies.

iii. A figure of merit health calculation

The figure of merit health value is based on the previous residual calculation and clustering. After selecting the cluster, the figure of merit value can simply be the arithmetic mean of the residual values in that cluster. For a multivariate case, the figure of merit value can simply be the summation of the residual mean values. These are the example metrics used in the case studies, but other health metrics can also be used.

4.3 Mathematical Description

The baseline data consist of an r by m feature matrix \( (F_{r \times m}) \), in which r is the number of samples and m is the number of selected features for monitoring the system or component. The baseline data is from the system or component in the nominal healthy condition and can be from a single unit or multiple units. Considering that wrapper and filter based feature selection methods require data from two or more classes [37], prior information is needed for selecting the condition monitoring features. The m selected features can be based on previous seeded
fault experiments on a similar system or component, prior experience, or the selection can be based on published work from the literature.

The baseline data can be further partitioned into a training data set consisting of \( n \) samples \((F_{n \times m})\) and a validation data set \((F_{r-n \times m})\). The training data set is used to train the residual processing model. Potential options for the residual processing model include a regression model, an auto-associative neural network, or a self-organizing map. Without any loss of generality, some example formulations are provided regarding the selection of the number of nodes and number of training samples needed for properly training the auto-associative neural network model. A similar approach can be followed when using the other algorithms for the residual processing.

![Auto-Associative Neural Network Diagram](image)

Figure 18: Auto-Associative Neural Network (9-5-3-5-9), \( \sigma \) is the tan sigmoid transfer function, \( L \) for linear transfer functions, \( f \) are inputs to the network and \( \hat{f} \) are outputs of the network.

An auto-associative neural network is a feedforward neural network which consists of 3 hidden layers and the network is trained with the same inputs and targets; an example structure
of an auto-associative neural network is provided in Figure 18. It effectively is learning the nonlinear correlation relationship among the feature values and consists of a mapping layer with \( M_1 \) nodes, a bottle-neck layer of \( b \) nodes, a de-mapping layer of \( M_2 \) nodes, and input and output layer of \( m \) nodes [92]. Assuming all the nodes in the network have a bias term, the total number of free parameters \( (N_p) \) is given by the expression in Equation 14.

\[
N_p = (m + b + 1)[M_1 + M_2] + m + b 
\]  

(14)

The number of free parameters should be much less than the number of data elements \((n \times m)\), this leads to the following relation for \( M_1 \) and \( M_2 \) provided in Equation 15. The number of nodes \( (b) \) in the bottle neck layer represents the intrinsic dimension of the features and is usually quite small. If one assumes that \( (b \ll m, n) \), then the previous expression can be approximated by a simpler form as shown in Equation 16. There is an additional requirement that the number of nodes in the mapping layer \( (M_1) \) and de-mapping layer \( (M_2) \) have to be greater than the number of nodes in the bottle neck layer; this is necessary for having enough nodes for extracting the nonlinear factors.

\[
M_1 + M_2 \ll \frac{m(n - b)}{m + b + 1} 
\]  

(15)

\[
M_1 + M_2 \ll n 
\]  

(16)

Upon selecting the appropriate number of nodes or parameters for the residual processing model, the parameters are calculated during the learning phase of the algorithm. Typically, a validation data set is then used to check whether the model can generalize and accurately predict the feature values for the additional baseline data set that it was not trained with. If the results from the validation step are not satisfactory, the model structure can be refined and retrained until a suitable model is obtained [93].
For each monitoring period \((T_i)\) \(w\) samples long, a feature matrix is calculated \((f_{ij} | i=1…w, j=1…m)\). A residual value \((P_{ij} | i=1…w, j=1…m)\) for each feature matrix is calculated using Equation 17, in which the predicted feature values \((\hat{f}_{ij} | i=1…w, j=1…m)\) are calculated from the trained data model. The residual values for the monitoring period are further processing using a clustering algorithm and a figure of merit health value calculation.

\[
P_{ij} = f_{ij} - \hat{f}_{ij} \tag{17}
\]

If one considers the previously calculated residual vectors as shown in Equation 18; the objective of clustering is to group the residual vectors. This maps the residual vectors into \(R\) groups \((G_k)\), in which each group represents a set of residual vectors as shown in Equation 19. For each cluster there would be \(L_k\) samples (for \(k=1…R\)), such that each \(G_k\) would consist of a matrix that has \(L_k\) rows and \(m\) columns.

\[
P = \{p(1),...,p(w)\}, \quad p(i) \in \Re^m \tag{18}
\]

\[
P \mapsto G_k \quad k = 1,…,R \tag{19}
\]

For this residual clustering method, usually the number of clusters would be selected to be two. This is based on the assumption that there is an operating regime in which the signature is more pronounced and the remaining operating conditions in which the signature is less pronounced. Focusing on a particular set of operating conditions is performed by the cluster selection routine. For many applications, there is some prior knowledge on the sign of the residual values and this can be used for cluster selection. For vibration based monitoring, vibration features higher than predicted are indicative of a faulty, while vibration features lower than normal are due to normal variation [89]. For the anemometer sensor case study, it is known that the dry friction whip failure mode of the anemometers cause a lagging wind speed reading and hence the negative residuals would offer more insight on the health condition of the
anemometers [86]. The initial step in the selection procedure is to calculate the mean residual values for each sensor from each cluster, as shown in Equation 20. A residual mean summation value can be formulated as shown in Equation 21 for each cluster.

\[ E_k(j) = \frac{1}{L_k} \sum_{i=1}^{L_k} G_k(i, j), \quad \text{for } k = 1: R, \quad \text{for } j = 1: m \]  

(20)

\[ S_k = \sum_{j=1}^{m} (E_k(j)), \quad \text{for } k = 1: R \]  

(21)

The cluster with the maximum or minimum \( S_k \) value for the \( R \) clusters is the selected cluster \( (G_{\text{selected}}) \). For a vibration based application, the maximum \( S_k \) value would be used; while the minimum \( S_k \) value is used in the cluster selection for the anemometer case study. The figure of merit health value is derived from the selected cluster \( G_{\text{selected}} \). An example formulation of the figure of merit health value is provided in Equation 22 – 24, in which it is simply the \( S_k \) value for the cluster. This example formulation is shown in which the positive residuals are used for the cluster selection. The figure of merit health value for the \( i^{th} \) monitoring period \( (h_i) \) is used for monitoring the engineering asset over time and represents the output of the residual clustering monitoring method.

\[ a = \arg \max_k (S_k) \]  

(22)

\[ G_{\text{selected}} = G_a \]  

(23)

\[ h_i = S_a \]  

(24)

4.4 General Framework and Variations

The general framework of the residual clustering approach for assessing the health of a system or component operating in multiple operating regimes is provided in Figure 19. The
initial step in the framework consists of a training stage for learning the characteristics of the machine or system from baseline data. Data pre-processing steps are a typical first step for preparing the data, in that outliers or incorrect context information can have a significant effect on the trained regression or residual processing model [94]. For highly sampled waveform signals such as vibration or current, various signal processing methods can be used to extract degradation features from the time, frequency, or time-frequency domain [12]. A data driven model is used to learn the correlation relationship between the extracted features or signals from the baseline data; this could consist of an auto-associative neural network model, or a regression model. Selecting the appropriate model form is discussed in more detail in the subsequent paragraphs about modeling options when using this proposed framework.

During the monitoring stage, measured signals from the monitored engineering asset go through the same data pre-processing and feature extraction steps that were performed on the baseline data. The trained model is then used to calculate residuals between the actual and predicted signal features. The trained model represents the nominal correlation relationships among the signal features. Deviations from this baseline behavior represent unexpected magnitude or patterns in the features that represent a degraded condition. Clustering the residual signal is based on the assumption that the signature is not present in all the operating regimes and hence more accurate health estimation can be obtained if one selects a particular regime subset for assessing the health condition. The clustering method incorporates prior knowledge for selecting the particular cluster and calculates a figure of merit health value using only samples from the selected cluster. Other health methods that are based on the comparison in the overlap or change in the data from the entire operating space could be diluted and provide a less incipient detection of the deteriorated health condition of the monitored system.
The outputs from the residual clustering monitoring method can also be used for designing future prognostics and health management systems (PHM). Instances when the figure of merit health value is above a predefined threshold can be accumulated over time for the monitored fleet of assets. This along with the clustering results can be used to highlight which operating regime was selected for these assets in which the health values were above the normal limits. The clustering results would provide information on which operating regime the degradation signature was most prevalent. A regime map could be formed to summarize these results and provide designers of future PHM systems insight on which monitoring regime to monitor and focus on. Simplified but accurate local health models could be used for future monitoring of these engineering assets after sufficient accumulation of evidence is provided by the residual clustering monitoring method. For applications such as machine tool or robots, in which a fixed cycle feature test could be used [77], the regime map would provide insight on which operating conditions to use during the routine program.
A visual diagram of the key processing steps and outputs of the residual clustering method is provided in Figure 20. The flow chart depicts some of the key steps in the transformation of the input signal into a figure of merit health value. The residual processing method shown in this example provides a case in which the residuals are quite large in certain instances, but also quite small in other instances. This example of a regime dependent signature is explored in more detail in the case studies on anemometer sensor fault detection and gearbox health monitoring. An example regime map is also provided, which highlights which operating regime is most conducive for detecting the degradation signature. The gearbox case study provides an
example calculation and interpretation of a regime map for detecting gear and shaft related problems.

Variations of the residual clustering monitoring framework include the selection of the residual processing model form, the clustering algorithm, and the figure of merit health value calculation. Whether the features have a linear or nonlinear correlation relationship is one of the key aspects to consider when selecting an auto-associative residual processing model. If prior work or knowledge suggests that the features are linearly correlated, the use of a linear principal component model would suffice. A nonlinear model such as auto-associative neural network model is more appropriate if the features exhibit nonlinear correlations [95]. For systems with insufficient prior knowledge, the nonlinear auto-associative model is a suitable option in that it would still generalize for the linear case. In applications in which there are not clearly defined relationships between input and output variables, the auto-associative modeling approach is a good option since it is simply learning the relationship in the input features.

However, if there is a known relationship between the input features and the operating variables, a linear regression model or a multilayer perceptron neural network would be an appropriate modeling option [96]. In addition, multiple regression or auto-associative models can be used in order to account for unit to unit variances among the monitored assets. The use of multiple models and a weighted residual calculation procedure is discussed in more detail in the anemometer sensor fault case study. In that example, the anemometers are installed on different towers and un-modeled sources of variation could arise due to the way the sensors were installed and also from any variation in the manufacturing of the anemometers [97].

Various algorithms exist for clustering, including hierarchical clustering methods, centroid based clustering methods, and distribution based clustering methods [98]. A more complete description of clustering algorithms can be found in [99]. The selection of the best clustering
algorithm typically requires prior knowledge on the data structure which is usually not known a priori. In this work, centroid based, hierarchical clustering, and density based clustering algorithms are investigated. Other options for clustering the residual signal could also be explored; however the initial results indicate similar health monitoring results when different clustering algorithms are used.

The figure of merit calculation presented in this work is based on the residual mean or summation value for the selected operating regime cluster. A single variable residual mean value is more appropriate for the sensor fault detection scenario in which one would like to know the health condition of each sensor. For examples in which a multivariate health metric is more appropriate, a residual mean summation value can be used. Examples of this residual mean summation value are provided for the gearbox monitoring example, in which vibration features from multiple accelerometers are fused for detecting an eccentric gear wheel and a shaft with imbalance. Other figure of merit health metrics could also be calculated after clustering; various statistics could be calculated from the residual signal besides the mean value currently proposed.

4.5 Simulated Correlation Example Data Set

4.5.1 Description of Data Set

A simulated data set with four signals was considered for a more thorough investigation of the performance of the residual processing algorithms. Four of the most common algorithms were considered, and consisted of an auto-associative neural network, a linear principal component analysis model, a linear regression model and a feed-forward neural network regression model with 1 hidden layer. The simulated data set is based on an example that was used by Hines et al. [16] but this example was modified in order to adjust the correlation level between the signals and also induce different types of signal anomalies.
Table 10: Test Number and Correlation Level

<table>
<thead>
<tr>
<th>Test Run Number</th>
<th>Signal 3 (a*t)</th>
<th>Correlation Between Signal 2 and Signal 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.24</td>
</tr>
<tr>
<td>3</td>
<td>0.15</td>
<td>0.39</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>0.44</td>
</tr>
<tr>
<td>5</td>
<td>0.25</td>
<td>0.58</td>
</tr>
<tr>
<td>6</td>
<td>0.30</td>
<td>0.63</td>
</tr>
<tr>
<td>7</td>
<td>0.35</td>
<td>0.69</td>
</tr>
<tr>
<td>8</td>
<td>0.40</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>0.45</td>
<td>0.77</td>
</tr>
<tr>
<td>10</td>
<td>0.50</td>
<td>0.78</td>
</tr>
<tr>
<td>11</td>
<td>0.55</td>
<td>0.79</td>
</tr>
<tr>
<td>12</td>
<td>0.60</td>
<td>0.83</td>
</tr>
<tr>
<td>13</td>
<td>0.65</td>
<td>0.83</td>
</tr>
<tr>
<td>14</td>
<td>0.70</td>
<td>0.83</td>
</tr>
<tr>
<td>15</td>
<td>0.75</td>
<td>0.86</td>
</tr>
<tr>
<td>16</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>17</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>18</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>19</td>
<td>0.95</td>
<td>0.88</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>0.89</td>
</tr>
</tbody>
</table>

The functional form of the signals are provided in Equations 12-15, in which signal 2 has an adjustable parameter that is varied at 20 different levels. By adjusting the argument in signal 2, the correlation between signal 2 and 3 can be adjusted. Effectively as the value (a) is increased
from 0.05 to 1, the correlation between signal 2 and 3 is increased. A listing of the test number, the signal 2 parameter value, and the correlation level between signal 2 and signal 3 are provided in Table 10. In addition to using the expressions listed in Equation 25-28, a Gaussian random noise is added to each signal and has a magnitude of 0.1 to provide a more challenging and realistic noise level for the simulated data set.

\[ \text{Signal}_1 = \sin(10 \times t) \]  
\[ \text{Signal}_2 = t, \text{ in which } t \text{ is from } -0.5 \text{ to } 0.5 \]  
\[ \text{Signal}_3 = \sin(a \times t) \]  
\[ \text{Signal}_4 = \sin(15 \times t) \]  

Four sets of data for each test run were generated, and consisted of both training and testing baseline data sets, an induced shift fault testing data set, and an induced fault drift testing data set. The training and testing baseline data set used the functional forms listed in Equation 25-28 and the only difference between the two sets would be the random noise that is added to each signal. The ability of the residual processing models to capture the baseline behavior of the signals for the different correlation level is evaluated by looking at the residual errors for the baseline testing data set. A simulated shift fault was induced by adjusting Signal 2 and suddenly changing the pattern from a linear increasing trend to a constant value of zero. This type of induced problem was carried out using the expression provided in Equation 29, in which the problem was induced halfway through the simulation. This simulated example could represent a sensor loose connection fault, where it suddenly is outputting just noise. This would effectively break-down the normal correlation relationship between the signals, since signal 2 and signal 3 are normally well correlated. It should be noted that this type of fault only induces a change in correlation between the signals, and that signal 2 is still within its normal range.
Another type of fault was also considered, in which signal 2 continued to increase over time, but at a faster rate than expected. This type of drift fault was induced by using the expression provided in Equation 30, in which the fault is also induced halfway through the simulation run. This drift type fault would induce a slight change in correlation between signal 2 and 3, since signal 2 is increasing at a greater rate than expected. However, since the values for signal 2 would also exceed the maximum values in the baseline state, it would also represent a case in which the signal values are outside the normal range. Thus the drift fault represents an example in which the system degradation is causing an abnormal increase in one of the sensor values.

\[
Signal_{2} = \begin{cases} 
    if \text{ sample } \leq 500 & s_2 = t \\
    if \text{ sample } > 500 & s_2 = 0 
\end{cases}
\]  

(29)

Regarding the algorithm settings, initial trails with the baseline data and the drift seeded fault indicated that the best performance was achieved with an auto-associative neural network that had 3 nodes in the bottleneck layer and 10 nodes in the mapping and de-mapping layer. In addition, the linear principal component analysis method had the best performance when 3 principal components were retained. For the neural network regression model, the best results were achieved with 3 nodes in the hidden layer. These settings for the four algorithms were kept constant for the complete set of simulated data sets with the varying correlation levels.

4.5.2 Sequential Probability Ratio Test

For determining whether the residuals for a given signal represented a normal state or anomalous state, a sequential probably ratio test is used. More specifics about how the sequential probably ratio test can be used for anomaly detection for machine condition
monitoring is provided in [100] and only the key aspects are reviewed here. The initial assumption regarding the SPRT test is that the residuals \((y_k)\) follow a normal distribution with zero mean and variance \((\sigma^2)\). Based on this normal distribution assumption, the likelihood of the sequence of residuals (N-samples) is given by Equation 31; the likelihood function is simply the product of the normal distributions.

\[
L(y_1, y_1, \ldots, y_n \mid H_o) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{N} y_k^2\right)
\]  

(31)

When considering the test data, the common assumption for the SPRT test is the residual mean \((M)\) is typically one standard deviation from the normal data residual mean. In addition, it is assumed for the test data that the variance in the residual values is the same as the normal data. With these assumptions, the likelihood is given by Equation 32 for the test data residual sequence.

\[
L(y_1, y_1, \ldots, y_n \mid H_t) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{k=1}^{N} y_k^2 - 2\sum_{k=1}^{N} My_k + \sum_{k=1}^{N} M\right)\right)
\]  

(32)

In order to calculate the SPRT test statistic, one needs to consider the likelihood ratio between the residual sequence for the test data and the residual sequence for the baseline data. Using the expressions in Equation 31 and Equation 32, the likelihood ratio is provided in Equation 33. The test statistic is shown in Equation 34, and is simply the natural logarithm of the likelihood ratio.

\[
L_n = \frac{L(y_1, y_1, \ldots, y_n \mid H_t)}{L(y_1, y_1, \ldots, y_n \mid H_o)} = \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{n} M(M - 2y_k)\right)
\]  

(33)

\[
SPRT_{\text{mean}} = -\frac{1}{2\sigma^2} \sum_{k=1}^{n} M(M - 2y_k)
\]  

(34)
For evaluating the hypothesis test, one would need to determine the acceptable type I and type II error for a given monitoring application. The type II error ($\beta$), would represent the probability of a missed detection by the monitoring system, while a type I error ($\alpha$) would be the probability of a false alarm. Based on the $\alpha$ and $\beta$ values, the system is considered an anomaly based on the inequality provided in Equation 35; effectively if the test statistic is greater than that limit an anomaly is flagged and is considered healthy otherwise. This SPRT test on the residuals is checked using both a positive residual mean test and also a negative residual mean test. In section 4.5.3, the SPRT test results are presented using the following convention: a zero if the system is considered normal, 1 for a positive residual anomaly, and -1 for a negative residual anomaly.

$$SPRT_{\text{mean}} \geq \ln\left((1 - \beta) / \alpha\right) \quad (35)$$

4.5.3 Residual Results for Different Correlation Levels

The initial evaluation of the trained residual processing algorithm is performed on an additional set of baseline data. If the residuals are high and the algorithm cannot model the baseline data well, then it will have a high propensity for false alarms. An example result from the auto-associative model is provided in Figure 21, in which the model can provide accurate predictions and low residuals for the additional baseline data set. This example is from test #20, in which the correlation between signal 2 and signal 3 is quite high at 0.89. However, even when the correlation was quite low (below 0.3), the auto-associative model still had a small amount of residual error. Section 4.5.4 has a more in depth set of results and discussion on how the auto-associative model had low residuals for the baseline data, regardless of the correlation level.
Of equal interest, was determining how well each of the four residual processing algorithms could accurately detect the shift fault. When this anomalous condition occurs, signal 2 no longer continues with its increasing trend but instead levels out at zero. The auto-associative mode could accurately detect this fault for the cases when the correlation level between signal 2 and signal 3 was quite high. An example result using the auto-associative model is provided in Figure 22, in which one can observe a noticeable difference between the predicted and actual values for signal 2 when the fault occurs. As one would expect, the predict sensor values are greater than the actual sensor values, this is because signal 2 is not following the increasing trend that signal 3 is experiencing.
To test the accuracy of each method, the residuals from sensor 2 were then input into the sequential probability ratio test (SPRT), which would provide a logic value for each sample on whether it was considered anomalous or not. Considering that the SPRT is a moving summation calculation, the first 750 samples were considered to be the normal state and the last 250 samples were considered the fault state. This ground truth could then be compared with the SPRT test on the residuals and thus the accuracy of each method could then be quantified. An example SPRT test result using the auto-associative model residuals is shown in Figure 23. This example result is also from test #20, in which the correlation between signal 2 and signal 3 is very high. The SPRT result indicates that the AANN model can detect the
problem right after sample 800; this is an accurate result with a late detection by only a few samples and no false alarms.

Figure 23: Sequential Probability Ratio Test for Residuals for Signal 2 - Shift Fault (Test 20)

The results for the shift related fault were quite interesting when the correlation between the signals was low. The auto-associative neural network model did not have enough sensitivity to detect this fault when the signals had low correlation. The residual processing results plotted in Figure 24, shows that the predicted and actual sensor values are in good agreement for signal 2, even when the shift fault is initiated. Effectively if there is low correlation in the signals, the AANN algorithm will have low sensitivity to correlation type changes, such as the shift type fault when one of the signals stops trending upward.
Unlike the induced shift fault, the drift fault would primarily cause signal 2 to have values outside its normal range and would only induce a minor change in the correlation between the signals. This induced drift fault is effectively increasing the magnitude of signal 2 at a rate faster than its normal linear rate. The auto-associative residual processing model could accurately detect the drift problem rather easily, as illustrated in Figure 25. The predicted values are expecting a linear increasing, and thus the residuals are quite high as signal 2 increases in a quadratic manner. Unlike the shift fault, the drift fault could be detected rather accurately using the AANN model, even when the signals had a small amount of correlation.
To provide some additional explanation on why the auto-associative model can still have accurate predicted sensor values when the correlation between the signals is low; some sample test inputs were provided to the AANN model for the test #1 case. The sample test inputs, predicted values, and residuals for test #1 are provided in Table 11, in which one can observe that the AANN output is only marginally influenced by the other signals when the correlation is low. This implies that the AANN model tends to believe the actual signal value instead of the other measured signals when the correlation is low. However, if the signals are well correlated, the other measured signals have more influence and weight for calculating the predicted sensor value. This aspect is illustrated by considering the AANN model for test #20, when signal 2 and signal 3 had a correlation of 0.89.
Table 11: Auto-Associative Model for Test # 1 (Low Correlation) – Example Outputs

<table>
<thead>
<tr>
<th>Signal 1</th>
<th>Signal 2</th>
<th>Signal 3</th>
<th>Signal 4</th>
<th>AANN Signal 2</th>
<th>Residual Signal 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.47</td>
<td>0.03</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.55</td>
<td>-0.05</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.41</td>
<td>0.09</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.55</td>
<td>-0.05</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.45</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0.43</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.38</td>
<td>0.12</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.35</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 12: Auto-Associative Model for Test # 20 (High Correlation) – Example Outputs

<table>
<thead>
<tr>
<th>Signal 1</th>
<th>Signal 2</th>
<th>Signal 3</th>
<th>Signal 4</th>
<th>AANN Signal 2</th>
<th>Residual Signal 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.41</td>
<td>0.09</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.26</td>
<td>0.24</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.65</td>
<td>-0.15</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.26</td>
<td>0.24</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.67</td>
<td>-0.17</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0.34</td>
<td>0.16</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.41</td>
<td>0.09</td>
</tr>
</tbody>
</table>

The test inputs and AANN outputs for this example are provided in Table 12, and one can observe that the other signals have a more significant affect on the predicted sensor response. In actuality, what this implies is that the AANN model will have low residuals for baseline data regardless of how well correlated the measured signals are. However, the low correlation could
potentially affect the algorithms sensitivity for detecting anomalous system health conditions and degradation.

4.5.4 Summarized Results for Correlation Data Set

The previous section highlighted some of the example results from the residual processing algorithms for the correlation data sets and in particular focused on the auto-associative neural network results. However, summarized plots on the results for the four different methods would also be beneficial and would provide some guidance on the advantages and disadvantages of each method. For the baseline data, the RMSE value was calculated for the auto-associative neural network, principal component analysis, neural network, and multi-linear regression models for each test run.

Figure 26: Root Mean Square Error Values with Different Correlation Levels for the Four Different Algorithms for the Baseline Data Set
The results are provided in Figure 26, in which the neural network, linear regression, and PCA models perform poorly when the correlation is low. The MLR, NN, and PCA models do have a noticeable increase in prediction accuracy when the signals are better correlated. This is in contrast to the AANN model which has low residuals and high prediction accuracy regardless of the correlation level. The AANN model has a lower RMSE value for all 20 test runs when compared with the other 3 methods.

The performance of each algorithm on the baseline data only provides a measure of the algorithms propensity for false alarms. The induced shift and drift faults would also check the algorithms monitoring accuracy. The results for the shift fault are plotted in Figure 27 for all four
algorithms. The top plot shows the auto-associative model result, in which the algorithm has an accuracy of only 0.75, until the correlation between the signals are above 0.8. After the correlation reaches that level, the algorithm has enough sensitivity to detect this type of problem. When the correlation is low, the AANN method does not have any false alarms but is not sensitive to the shift fault.

The PCA, neural network, and linear regression models have a high false alarm rate and low sensitivity when the correlation between the signals is in the range of 0.1 to 0.7. The MLR, neural network, and PCA models have high residual error for the baseline data when the correlation is low, and thus have a high false alarm rate under these conditions. For this example shift fault, the AANN method had similar detection sensitivity when the correlation was high but a lower false alarm rate when the correlation was low. Thus the AANN method had the best accuracy overall for this scenario.

The same type of summarized results for the simulated drift fault is presented in Figure 28. In the top plot, one can observe that the AANN is quite accurate for the simulated drift fault and that the accuracy is not influenced by the level of correlation between the signals. The PCA, MLR, and neural network models are inaccurate when the correlation is low, but continue to improve as the correlation between the signals is increased. The PCA, MLR, and neural network models have a high false alarm rate and low accuracy in general if the signals are not well correlated regardless of the type of fault. However, the AANN model has a low false alarm rate and a detection sensitivity that is dependent on the fault type and also the level of correlation in the signals.

For magnitude changes, the AANN model has high detection sensitivity, even when the correlation between the signals is low. However, for detecting changes in the correlation structure, the AANN sensitivity is a function of the amount of correlation in the measured
signals. In a general sense, the results go along with the assumption that the signals should be well correlated when considering the residual clustering health monitoring approach. However, the AANN model can still be used if the system failure modes would cause changes in the measured signals magnitude as opposed to changes in the correlation structure.

![Accuracy of Detection for Simulated Drift Example](image)

Figure 28: Accuracy of Detection for Simulated Drift Example for All Four Methods

4.6 Comparison of Residual Processing Methods

4.6.1 Description of Data Set

Understanding the performance of the different algorithms for providing predicted values and calculating residuals is imperative, considering that the residual processing module is a key aspect of the proposed health monitoring method. As an extension of the results from the
simulated data sets with different correlation levels, further evaluation of the residual processing methods is conducted with actual field data from a wind energy application using anemometer wind speed sensors. The data set consists of wind speed sensors at different height locations along a metrological tower, in which there is a non-linear correlation relationship between the different wind speed measurements. The data set is described in more detail in Chapter 5 for the sensor health monitoring case study, and only a portion of the baseline data is used for this residual processing algorithm evaluation.

An auto-associative neural network model, a linear regression model, a neural network model, and a principal component based model are evaluated for this wind energy application in terms of their ability to generalize and accurately predict the wind speed sensor values. The scenario is designed by training each algorithm with a set of baseline data from a particular unit (metrological tower) and then testing the algorithms performance with a data set from another tower. There are 7 towers provided in the baseline data set and in this example unit 4 was selected for training and unit 5 was selected for testing; these two units both have the same tower configuration with 4 anemometers at the same height location. It was important to understand whether each of the four algorithms could accurately capture the correlation relationship among the wind speed sensors and were not over-fitting and memorizing the training data set. By training and testing the algorithm with two separate units, this would provide a good measure of whether the algorithms could generalize and could properly model the correlation relationship in the four wind speed sensors.

In addition to evaluating how well the residual processing algorithms could generalize, it was also important to consider the variability in each algorithm. A set of 50 runs was conducted in which 70% of the samples from unit 4 were selected for training and also 70% of the samples from unit 5 were also randomly selected for testing. Thus for each run, a different random set of samples is selected and the accuracy of each algorithm was evaluated in terms of the root-
mean square error (RMSE) for each run. The results from the 50 runs for each algorithm can then be summarized as a mean and standard deviation in terms of the RMSE values, which would allow one to rank the algorithms in terms of accuracy and also variability.

4.6.2 Residual Processing Algorithms and Settings

For a fair comparison between each of the algorithms, the same input signals were used. In this case study, there were 4 anemometer wind speed sensors on the tower, and a total of 12 parameters since a maximum, minimum, and mean value statistic value was provided for each sensor. For the auto-associative neural network (AANN) and principal component analysis (PCA) algorithms, both of these algorithms are trained with the 12 input parameters. Both the AANN and PCA methods provide a predicted value for each of the 12 inputs for each test data sample. The multiple linear regression model (MLR) and the neural network method would require multiple models in order to have predicted values for each of the 12 parameters. Training 12 different models would be too cumbersome. It was decided to evaluate the residual processing methods by considering the predicted and residual value for the mean wind speed sensor value for the anemometer at the greatest height on the tower. For the regression and neural network methods, this implied training the algorithm with 11 of the parameters as inputs and the mean wind speed value for the highest anemometer as the target variable.

In addition to using the same input parameters for each algorithm, it was important to conduct the comparison using the best settings for each algorithm. For the auto-associative neural network, the most influential parameter is the number of nodes in the bottleneck layer. The residual processing results for the AANN were compared using a different number of nodes in the bottleneck layer (ranging from 1-4). The results in terms of their RMSE values are presented in Figure 29, in which one can observe that the number of bottleneck nodes has a significant effect on the prediction accuracy of the model. The best prediction performance for the AANN model was achieved with 4 nodes in the bottleneck layer and better performance in
general was achieved as the number of nodes was increased. Using too few of nodes can lead to poor reconstruction and high prediction error, however the number of nodes in the bottleneck layer should also be less than the number of nodes in the mapping and de-mapping layer. Considering the results from adjusting the number of bottleneck nodes, the final AANN configuration consisted of 7 nodes in the mapping and de-mapping layer and 4 nodes in the bottleneck layer. This AANN structure of 12-7-4-7-12 was then trained and tested during the 50 runs and compared with the other algorithms.

Figure 29: Residual Processing Results for AANN Using Different Number of Bottleneck Nodes

Figure 29: Residual Processing Results for AANN Using Different Number of Bottleneck Nodes
Figure 30: Neural Network Results with Different Number of Hidden Layer Nodes

The neural network regression model prediction accuracy is also dependent on the number of nodes used in the hidden layer. For the neural network model, the configuration consisted of an input layer, a single hidden layer, and an output layer, in which the first transfer function consisted of a tan-sigmoid function and a linear transfer function was used between the hidden and output layers. The only adjustable parameter for the neural network model is the number of nodes in the hidden layer, and this was varied from 1 to 10 in order to determine which configuration provided the best prediction accuracy. The prediction results for the neural network model as a function of the number of nodes in the hidden layer are provided in Figure 30. The results indicate that the neural network prediction results are quite consistent and the numbers of nodes in the hidden layer only have a marginal influence on the prediction accuracy.
The prediction accuracy was the highest when 3 nodes in the hidden layer were used and this setting was used for comparing the neural network model to the other 3 algorithms.

For the principal component based residual processing method, the best performance was achieved when 4 principal components were used. This implies that the data is best represented with four linear components, which also goes along with the four non-linear components that provided the best results for the AANN method. Thus, for the algorithm comparison study, four components were used for the PCA model.

For the multiple linear regression (MLR) method, potential variants of the model form could have considered, such as the inclusion of quadratic or cubic terms or interaction variables. However, this would have provided a different set of inputs for this particular method when compared with the other 3 algorithms. Using a different set of inputs would make it more difficult to draw conclusions on which method provided the best performance. Considering this aspect, the linear regression model consisted of only the 12 coefficients (off-set term and coefficients for the 11 inputs) and interaction or higher order terms were not used.

4.6.3 Summary of Residual Processing Results

After selecting the settings that provided the best prediction performance for each algorithm, the RMSE prediction accuracy was calculated for a set of 50 runs for each method using the wind speed sensor data set. A summary of the prediction RMSE results are provided in Figure 31, in which the top plot shows a box plot for each method and the bottom plot shows the RMSE mean value. In general, the auto-associative neural network model had the best prediction accuracy of the four methods, with the lowest RMSE mean value. In addition, the AANN method clearly outperformed its linear counterpart, the PCA residual processing method. The AANN method had a mean RMSE value of 0.16 compared to a RMSE value of 0.23 for the PCA method. The neural network based regression model and the linear regression model had
similar performance in terms of the RMSE average value. The neural network model had more variation than the linear regression model, in which a few runs had RMSE values below 0.2 but also had a few runs with RMSE values above 0.24.

Despite having the best prediction performance in terms of the RMSE average value for the 50 runs, the AANN method had more variation in the residual processing results than the other 3 algorithms. For example, the AANN method had a few runs in which the RMSE value was above 0.25. However, the vast majority of the runs had low RMSE values and good prediction performance. Table 13 summarizes the results from each method in terms of their mean and standard deviation. The auto-associative model had the largest standard deviation value despite having the lowest RMSE average value. The linear models such as PCA or linear
regression had less variation but also lower accuracy. The neural network model also had a considerable amount of variation and less prediction accuracy than the auto-associative neural network model.

Table 13: Residual Processing Result Summary (Average and Standard Deviation)

<table>
<thead>
<tr>
<th></th>
<th>AANN (BN-4)</th>
<th>PCA (4)</th>
<th>NN (3)</th>
<th>MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>0.1656</td>
<td>0.2265</td>
<td>0.2149</td>
<td>0.2143</td>
</tr>
<tr>
<td><strong>STD</strong></td>
<td>0.0203</td>
<td>0.0077</td>
<td>0.0120</td>
<td>0.0034</td>
</tr>
</tbody>
</table>

4.6.4 Discussion

Evaluating the four different residual processing algorithms for the wind speed sensor data set provided some interesting insights on the performance of each algorithm. In this data set, there was an expected non-linear correlation relationship between the wind speed measurements at the different heights, since previous research has suggested using a power or logarithmic relationship to model the wind speed and height relationship [101]. In that sense, it was not surprising that the auto-associative neural network model outperformed the linear principal component model in terms of prediction accuracy. Effectively, the auto-associative model is a non-linear version of the traditional linear PCA model and is thus more suited for modeling non-linear correlation relationships in the measured signals.

The neural network and linear regression models are less suited for data sets in which there is not a clearly defined target variable and input variables. In addition to the neural network and the regression model providing lower prediction accuracy than the auto-associative model, they also would be more cumbersome to use in practice. In this application, one would want to calculate predicted values and residuals for all 4 wind speed sensors and this would require training multiple linear regression models or neural network models.
Although the auto-associative model provided the best prediction accuracy, it also had the most variability in terms of its prediction performance. Training a 3 layer feed-forward neural network has more variation from run to run when compared with the matrix operations need for finding the eigenvectors and eigenvalues used by the PCA model. The only variation for each run for the PCA model comes from using a random set of samples, in which 70% of the data from unit 4 is selected for training and 70% of the data from unit 5 is used for testing. These results suggest that an auto-associative model is more accurate if it is believed that the sensors or features have a non-linear correlation relationship. However, in order to reduce the risk of using an AANN that is less suitable, it is best to train multiple AANN models and select the model that has the best prediction performance.
CHAPTER 5: CONTINUOUS OPERATING CONDITIONS – ANEMOMETER SENSOR
HEALTH MONITORING CASE STUDY

The proposed methodology is demonstrated in this first case study, which is from the 2011 Prognostics and Health Management Society Data Challenge. Using the residual clustering approach resulted in a first place finish among 24 total participants from academia and industry. The data set provided by the PHM Society in this competition is used to show the effectiveness of the residual clustering approach for assessing the health condition of a multi-regime system. The method and results presented here is based on the authors recently published journal paper in the International Journal of Prognostics and Health Management [97].

5.1 Introduction

One of the fundamental requirements for data interpretation, model development, and system monitoring is the need to have properly working and calibrated sensory data [102]. Considering the importance of properly working sensors, there is considerable research in the area of sensor fault detection and diagnosis with a diverse set of applications ranging from automotive [103], aerospace [104], and nuclear power plants [105]. The wind energy in particular, is quite reliant on obtaining accurate sensor measurements of wind speed, since this ultimately is one of the inputs used to estimate the energy production for a given site [106]. During feasibility studies of potential wind turbine sites, anemometers placed on meteorological towers are used to provide information on the long term wind speed characteristics. Historical wind speed data is one of the inputs provided to sophisticated meteorological models that provide an estimation of the energy production for a given site. Errors in the wind speed measurements can have significant effects on the estimated energy production, which impacts the return on investment for a given site or whether the site is financed [107].
Recent work in the area of anemometer fault detection includes the work by Kusiak, Zheng, and Zhang [108], which propose a virtual sensor method using a multilayer perceptron neural network. This study also discusses the use of a wavelet de-noising method for data pre-processing and a control chart based on the residuals calculated from the predicted and measured wind speed. A more classical statistical approach was discussed in the work by Beltran, Llombart, & Guerrero [109], in which a metric was derived from the difference in the 10 minute wind speed average data between two anemometers in close proximity. In addition to this prior work, a recent study by Clark, Clay, Goglia, Hoopes, Jacobs, and Smith [86] was done to investigate the root cause of NRG #40 anemometers reading slower than the actual wind speed. The study discussed statistical metrics, signatures of anemometers with excess measurement error, calibration methods, and a physical explanation of the potential failure mode that was believed to be the cause of the sensor measurement error.

5.2 Data Description

The 2011 Prognostics and Health Management Society [110] presented a data challenge problem dealing with this increasingly important topic of anemometer fault detection. Two different types of data sets titled the “paired data set” and the “shear data set” was provided for developing and evaluating anemometer fault detection algorithms; the results from this case study are for the shear data set. The shear data set consisted of data collected from three or four anemometers at different heights on the meteorological tower. Statistics from the wind speed sensors, a wind direction measurement, and ambient temperature reading were provided. The statistics were calculated from a 10 minute time period and consisted of the mean, standard deviation, maximum, and minimum for each parameter. Height information was also provided for each anemometer. In total, 28 or 23 parameters were provided in each data file, the difference in the number of parameters is due to certain sites only having 3 anemometers instead of 4 anemometers. A total of 7 Training data sets that comprised of 25
days worth of data were provided for the shear data set. The training data sets provided data from anemometers in a healthy condition. Test files were also provided in which the health condition of the anemometers were blind to the contest participants. The test files consisted of 225 files, with each file representing 5 days worth of data. The objective in the shear data set was to determine whether the set of anemometers were in a nominal healthy state or at least one of the anemometers had a fault and were exhibiting excessive measurement error [110].

5.3 Data Processing Methods

A flow chart of the health assessment algorithm for the anemometer shear data is provided in Figure 32. The algorithm used in this study has a training and monitoring phase, in which the training phase is developed using anemometer data from a nominal healthy state. The initial step in the training process is to perform data filtering. The data filtering step is designed to remove instances in which icing could occur as well as to remove other data samples in which there could be erroneous readings in wind speed, temperature, or other sensor measurements. The data normalization step is a specific step designed for the shear data and is based on the wind profile power law [101]. This normalization procedure uses the power law equation to place the wind speed measurements at a common reference height.

The data normalization step reduces the variation due to elevation; however an auto-associative neural network is used to further model the relationship and correlation structure between the anemometer wind speed statistics. In this study, multiple baseline data sets were available for model training and this provided the opportunity to have multiple auto-associative neural network models. The training of the auto-associative neural network models completes the training process and the algorithm can then be deployed in its monitoring phase.

In the shear data set, a given monitored shear data set file consisted of either 3 or 4 anemometers and each data file comprised of 720 samples and a duration of 5 days. Thus,
the data processing and health decision is performed on data from that 5 day period for a monitored shear anemometer set. The initial step for the monitored anemometers consist of performing the same data filtering and normalization method that was used for the training set.

A weighted residual calculation is performed using the auto-associative neural network models; a weighted approach is used in order to favor results from training models that more accurately predict the anemometer wind speed statistics. The residuals for the mean wind speed for each anemometer are then further processed in a k-means figure of merit calculation. The motivation for using a k-means clustering method is that prior literature suggested that the anemometers display a bimodal behavior in one of its failure modes and experience slowdown for a certain range of directions and wind speeds [111]. Thus, the residuals might be quite small in a particular speed or direction regime and could be potentially quite higher in a different

Figure 32: Anemometer Health Assessment Algorithm Flow Chart
regime subset. A figure of merit calculation is performed for each anemometer, and a decision on the health status for each anemometer is made based on whether the figure of merit value exceeds the threshold.

5.3.1 Data Filtering

The filtering routine is done to remove samples in which icing could be occurring and also for filtering out samples in which there are erroneous sensor values. For removing instances in which icing is occurring, there is a variety of parameters that could be used to infer this condition; the wind speed direction standard deviation statistic in particular is quite useful for filtering out icing events. Considering that various statistics are calculated for each 10 minute data block, a value of zero in the wind speed direction standard deviation would imply that there is no variation in the wind speed direction for a 10 minute time period. Physically this is not possible and this condition of no variation in the wind speed direction is one of the key parameters that can be used for filtering out samples in which icing could occur.

Table 14: Filter Settings for Shear Data Set

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Filter Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anemometer Mean 1</td>
<td>0.5 m/s – 26 m/s</td>
</tr>
<tr>
<td>Anemometer Mean 2</td>
<td>0.5 m/s – 26 m/s</td>
</tr>
<tr>
<td>Anemometer Mean 3</td>
<td>0.5 m/s – 26 m/s</td>
</tr>
<tr>
<td>Ambient Temperature Mean</td>
<td>-40°C - 120°C</td>
</tr>
<tr>
<td>Wind Direction Mean</td>
<td>Greater than 50 degrees</td>
</tr>
<tr>
<td>Wind Direction standard deviation</td>
<td>Greater than 0 degrees</td>
</tr>
</tbody>
</table>

The filtering settings used for the shear data set are provided in Table 14. For a given sample for the shear data set, it would have to satisfy all the listed ranges shown for the wind speed means, ambient temperature, wind direction mean, and wind direction standard
deviation. It was observed that instances in which the wind direction were quite low resulted in more sudden changes in wind speed mean values. This resulted in larger differences between anemometer wind speed readings during these more abrupt changes. Considering this aspect, the filtering routine includes logic for the wind direction mean parameter to remove these samples in which the wind direction is below 50 degrees. It was also observed in the training data sets that the initial samples in each data file contained erroneous sensor values, thus the filtering routine also removed the first 20 samples.

Figure 33: Example Filtering Results: Training Data Set 6 - Shear Data Set

An example of how the filtering removes potentially icing events and erroneous data samples is shown in Figure 33. This example is from a shear training data set in which all 3 anemometer sensors are in a nominal healthy condition; however, there is still a substantial amount of samples highlighted in green that are filtered out. The top plot in Figure 33 shows that the wind speed mean can have extreme high or low values, as indicated by the outlier value
of approximately 100m/s. The middle plot shows the second anemometer wind speed mean reading and one can observe that there are several instances in which both anemometers are reading at or near zero. These near zero readings are likely due to icing. The wind direction standard deviation is shown in the bottom most plot and this parameter is also zero during these suspected icing samples. This example highlights that the filtering algorithm provides an adequate detection of icing and outlier samples.

5.3.2 Data Normalization

The data pre-processing for the shear data set includes an additional step of data normalization in order to compare the wind speed measurements at a common reference height. In prior work in the literature, the wind speed profile has been modeled by a logarithmic relationship and also by a power law model [101]. The use of the logarithmic equation includes an additional aerodynamic surface roughness parameter that depends on the site location; this was not provided in this study and thus only the power law equation was used for data normalization. The power law equation is described by Equation 36, in which $u_1$ and $z_1$ are the wind speed and height at a known reference point and $u_2$ and $z_2$ are the wind speed and height at a location of interest. The exponent $p$ is a constant that is based on prior experimental studies and regression fitting; a value of 1/7 is a common value for this constant and one that is used in this study [112]. For data normalization, each wind speed measurement is corrected to a height of 49 m. In Equation 36, this would imply that $z_2$ is assigned a value of 49, while $u_1$ and $z_1$ are the known wind speed measurement and elevation for a given anemometer and $u_2$ is the corrected wind speed measurement at a height of 49 meters.

$$\frac{u_2}{u_1} = \left(\frac{z_2}{z_1}\right)^p \quad (36)$$
Figure 34: Shear Training Set 1 - Raw Wind Speed Mean Signals

Figure 35: Shear Training Set 1 - Normalized Wind Speed Mean Signals
An example of the normalization process is illustrated in Figure 34 and Figure 35. Figure 34 is from the first shear training data set and is comparing the wind speed for anemometers 1 and 4. With regards to the numbering convention, anemometers 1-4 are sorted from the highest to the lowest height and in this example have a height of 59, 50, 30, and 10 meters respectively. As one can observe, there is significant differences in the raw wind speed values for anemometer 1 and 4; these two anemometers have the largest difference in elevation. Figure 35 shows the normalized wind speed mean values for anemometer 1 and 4 from the same training data set. From visual observation, one can observe that the differences in the normalized wind speed values are lower when compared with the raw data.

In order to quantify the differences in the wind speed measurements, the Root Mean Square Error (RMSE) is shown in each plot. The RMSE can be calculated between two anemometers by using Equation 37, in which $N$ is the number of samples in the data file and $u_1$ and $u_2$ are the wind speed mean values for the two anemometers considered in the calculation [113]. The RMSE value for the normalized wind speed data in Figure 35 is much smaller than the RMSE value for the raw data in Figure 34; indicating that the normalization provided some measure of correction for the different elevations.

$$\text{RMSE} = \left( \frac{1}{N} \sum_{i=1}^{N} (u_1(i) - u_2(i))^2 \right)^{1/2} \quad (37)$$

5.3.3 Residual Processing

When a dynamic model of the system is not available a priori, the use of data driven health monitoring algorithms becomes a suitable alternative for monitoring the system health state [114]. Although there are various regression or distance from normal based metrics that are available, the use of an auto-associative neural network (AANN) has some intriguing characteristics that make it particular suitable for this application. Its ability to learn nonlinear
correlation relationships and calculate residual values for each sensor provides a means to calculate a system health value. In addition, contribution plots for each sensor can also be used to provide diagnostic information [69]. These attractive attributes of an auto-associative neural network have seen its usage for health monitoring span a diverse set of applications; from diesel engines [70] sensor health diagnostics and calibration [71], to commercial aircraft engines [75].

The theory and mathematics for the AANN were first described by Kramer [92] and this method is effectively a way to perform nonlinear principal component analysis. Although principal component analysis (PCA) has been used in a variety of applications for process monitoring by using Hotellings’ \( T^2 \) statistic and the residual square prediction error statistic (SPE); its assumption of the signals being linearly correlated is not satisfied in many engineering systems [72]. An auto-associative neural network provides a similar framework, but has the ability to learn the nonlinear correlation relationship among sensor variables. In this application, the underlying correlation relationship between the shear anemometer sensors is potentially nonlinear; this is suggested by the power or logarithmic equations used to relate wind speed and elevation. The auto-associative neural network is applied after data normalization is done to correct for the wind speed height. However, the difference in the anemometer wind speed values in Figure 35 implies that the underlying relationship is not completely described by the power law. An auto-associative neural network can be used to further learn the sensor correlation relationship.

The AANN model structure consist of 5 layers, an input layer, mapping layer, bottle neck layer, de-mapping layer, and an output layer as shown in Figure 18. One of its interesting aspects is that the network is trained with the same inputs and targets and thus the network is performing an identity mapping in which the output layer is providing an approximation of the inputs. The structure of the network used in this study follows the suggested configuration
provided by Kramer [92] and consists of 4 transfer functions. In sequential order, they consist of a tan-sigmoid transfer function, a linear transfer function, a tan-sigmoid transfer function, and a linear transfer function.

Although the network structure uses the same transfer functions, there were some minor differences in the auto-associative neural network models for the 3 or 4 sensor shear anemometer configurations. The inputs for the network consisted of the wind speed mean, maximum, and minimum values for each anemometer; this provided 12 and 9 inputs for the 3 and 4 shear anemometer configurations respectively. The structure of the AANN model used in this study was configured so that the numbers of nodes in the mapping layer were the same as the number of nodes in the de-mapping layer. The number of mapping and de-mapping nodes consisted of 5 and 7 for the 3 and 4 anemometer configurations. In both configurations, the bottle neck layer consisted of 3 nodes. The number of bottleneck nodes represents the intrinsic dimension of the data in a similar sense to the number of principal components retained in linear PCA [92]. The number of training samples (n) consisted of approximately 2000 samples for training the seven auto-associative neural network models. The number of training samples varied for each data set due to the data filtering routine removing a different number of samples for each data set. The de-mapping and mapping nodes summed to 14 which is much less than the number of training samples (2000). This satisfies the previously mentioned guideline for the AANN structure suggested in Equation 3.
As an additional extension of using the AANN models for anemometer health assessment, it was postulated that it might be advantageous to have an ensemble of training models. This could provide a way of giving more weight to training models that provide a more accurate sensor prediction for a given anemometer shear test data file. The rationale for considering this aspect is that there are several un-modeled sources of variation. Variation due to manufacturing, site topography, and installation, could potentially impact the AANN model accuracy. Using a weighted approach allows one to weight training models that might more closely represent the test data set. This can reduce variances due to other factors and allows one to assume that the deviation from the model is due to anomalous anemometer sensor behavior. A conceptual diagram of the weighted residual approach is highlighted in Figure 36 and the details of the calculation procedure are further described in this section.

The baseline files provided for the shear anemometer data set consist of two training baseline files for the three anemometer configuration and five training baseline files for the 4
anemometer configuration. The weighted AANN residual approach uses 7 trained AANN models for each of the baseline files, with 2 being assigned to the three anemometer configuration and 5 assigned to the 4 anemometer configuration. For a given test file, the residuals for each anemometer sensor statistic would be calculated for each model that matched the anemometer configuration for a given test file.

The residuals for each sensor statistic are weighted by a weight vector that is calculated from the sum of square error value as shown in Equation 38-39. In this calculation, $SSE_k$ is the sum of square error for the $k^{th}$ AANN model, and $P_{ijk}$ is the residual based on the predicted AANN value and the actual sensor statistic value for the $i^{th}$ data sample, the $j^{th}$ sensor, and the $k^{th}$ AANN model. In addition, $N$ is the number of samples in the data file, and $m$ and $r$ is the number of input parameters and AANN models respectively. The weight for each model is calculated by taking the models $SSE_k$ value and dividing that quantity by the summation of all the reciprocal $SSE_k$ values as shown in Equation 39. The weighted residual $(\hat{P}_{ij})$ is then calculated by taking the weights for each model multiplied by the residuals as indicated in Equation 40. This provides a residual value for each sensor statistic that includes aspects from each training model, but provides more weight for training models that more closely match the test data set.

\[
SSE_k = \sum_{j=1}^{m} \sum_{i=1}^{N} (P_{ijk})^2 \quad \text{for } k = 1 \text{ to } r \tag{38}
\]

\[
w_k = \frac{(SSE_k)^{-1}}{\sum_{k=1}^{r} \left( \frac{1}{SSE_k} \right)} \tag{39}
\]

\[
\hat{P}_{ij} = \sum_{k=1}^{r} P_{ijk} w_k \quad \text{for } i = 1 \text{ to } N \text{ and } j = 1 \text{ to } m \tag{40}
\]
In order to evaluate the generalization of the AANN models and the weighted residual processing, the baseline data sets were randomly divided into a training set and a calibration set, in which the training set consisted of 70% of the available samples in a given baseline data file. An example of how well the predicted sensor statistic values match the actual values are shown in Figure 37 for the first shear baseline data file. In this example, the blue curve represents the weighted predicted value from the AANN models and the red samples are the actual wind speed mean values. A measure of the model fit can be assessed by the root mean square error value (RMSE). In this example, the RMSE value is significantly lower when the AANN models are used as opposed to the results in Figure 35 that were obtained with only data pre-processing and normalization.
The trained AANN models and weighted residual processing method were then applied to the shear test files; example plots are shown in Figure 38 and Figure 39. In Figure 38, the results are for the first shear testing in which the predicted anemometer 1 wind speed mean and the actual anemometer 1 wind speed value are shown. Notice that the predicted and actual values match for the entire data set and the RMSE value is quite low. This is an example file in which the anemometers were considered to be in a healthy state. An example of the anemometers in a degraded state is provided in Figure 39. In this example, there is a noticeable difference in the predicted and measured anemometer wind speed mean values for the second anemometer. The RMSE value for this case is also quite high. The bimodal fault signature is also observed, since the sensor is only lagging for a portion of the data samples. This highlights the motivation for clustering the residual signal, since the signature is only present for a particular subset of the operating conditions.
5.3.4 Residual Clustering and Anemometer Health Value

There are a variety of techniques used in data mining and artificial intelligence for clustering and density estimation [115]. In this study, the k-means algorithm was used for partitioning the residual wind speed mean values into two clusters. Density estimation using Gaussian mixture modeling was originally considered; however, the computation time became burdensome given the number of data files that had to be processed, and the k-means algorithm provided a more efficient way of determining the data clusters. The k-means clustering algorithm aims to partition the data set into a set of R clusters, where R is the number of clusters specified and its objective function is to minimize the within cluster sum of squares [116]. The algorithm is iterative in nature, in that it is initialized with a random set of centroids and through the iteration process, updates the center locations in order to reduce the within cluster sum of squares.
distance. The interested reader is referred to the work by Hartigan and Wong [117] for a more
detailed description of the k-means algorithm.

Although the k-means clustering does not guarantee a global solution, 5 replications are
used in this study in order to avoid selecting a local minimum. The mean value is calculated in
each cluster and the minimum value of the two clusters is stored and denoted as the figure of
merit value. There is an additional logic rule to prevent a small sample cluster from being
included. If the sample size of one of the two clusters is below 60 samples, the mean of the
other cluster is stored as the figure of merit value. A small cluster could be due to a small
amount of outlier samples that potentially made it through the data filtering screening. In this
application, the figure of merit health calculation is based on the residual values for each
individual sensor and not a multivariate summation metric. The rationale for selecting this figure
of merit calculation is that this would provide a health value for each sensor and allow one to
determine which anemometer on the meteorological tower is not working properly. The
motivation for selecting the cluster with the minimum mean value is based on the prior literature
that suggest that a degraded anemometer would be reading slower than normal [86].

An example of the k-means clustering result is provided in Figure 40. This result is from the
residual wind speed signal for the system in a degraded health state. The histogram of the
residual wind speed shows a bi-modal distribution in the top plot; the clustering result in the
bottom graph indicates the two clusters that were determined using k-means clustering.
Considering that the figure of merit value is based on the mean value of the smaller of the two
clusters, k-means clustering provided a way of focusing on the samples when the anemometer
is lagging.
To further illustrate how the residual clustering approach can offer information on which operating regime the signature is most noticeable, histograms of the wind speed and direction are provided in Figure 41. In this example, the bottom graphs show the wind speed and direction distribution for the selected cluster. The signature appears to be prevalent for only a subset of the wind direction operating regime (around 100 degrees). If one were to calculate statistics on the entire distribution without clustering, the algorithm would be less sensitive to the fault signature.
5.4 Results

The previous section described how the figure of merit health values was processed; however the algorithm ultimately has to provide a decision statement on the health condition of each anemometer. This required setting thresholds for the figure of merit values. The literature suggests that an anemometer that is experiencing an increased level of friction and reading slower than normal could have an error of 1.5% to 3.0% and sometimes as high of a bias as 6% [111]. The thresholds were based on selecting a value within that error range. The figure of merit thresholds for the shear anemometers were set at -0.35 m/s for the three highest
anemometers and a threshold of -0.5m/s for the anemometer at the lowest elevation. The anemometer at the lowest elevation was set with a more conservative threshold since it was believed that the AANN predicted values had more error for this anemometer. One should note that many of the shear files only had 3 anemometers, so in many instances only the first 3 thresholds are used.

Considering that a fault is based on a lagging anemometer, a fault is declared if any of the figure of merit values are below its threshold and healthy otherwise. An example result for the figure of merit values is provided in Figure 42; this result is for the first anemometer for the shear data set. In this example, one can observe that the majority of the files are above the threshold. In total, 62 of the 255 shear anemometers were considered to be in a faulty state. Using the residual clustering algorithm resulted in a score of 421, the highest score among all 24 participants [110].

![Figure 42: Figure of Merit Results for Shear Data Set](image)
5.5 Discussion of Results

This case study applied the residual clustering health assessment methodology for assessing the health condition of anemometers. The methodology consisted of a series of algorithmic processing steps from data filtering, to a residual calculation, to a k-means figure of merit health value. The results from the case study were quite encouraging considering the method outperformed other techniques in the data challenge competition. In a general sense, this algorithm could be applied to many other sensor health monitoring applications. In particular, the use of auto-associative neural networks and the k-means clustering approach would be advantageous when redundant sensors are not available and the sensors have a regime dependent signature. The weighted residual processing using multiple baseline models is also useful for handling unit to unit variances since it weights training models that more accurately match the monitored unit.

Reconfiguring the algorithm for other applications is further explored in the second case study in Section 7, in which the system consists of a parallel stage gearbox with discrete operating regimes. Variations on the residual clustering algorithm are also considered by comparing the results for both a centroid based and density based clustering method. The second case study also benchmarks the proposed method with a distribution overlap health assessment method that also has been proposed for health monitoring of systems operating in multiple regimes. In addition, the notion of a regime monitoring map is further examined for the gearbox application, in which speed and load regime maps are generated for different gear and shaft failure modes.
CHAPTER 6: DISCRETE SET OF OPERATING CONDITIONS – GEARBOX
CONDITION MONITORING CASE STUDY

The residual clustering methodology is further demonstrated in this second case study, which is based on a similar gearbox data set that was used in the 2009 Prognostics and Health Management Society Data Challenge. It should be noted that the author’s winning algorithm for the 2009 data challenge was for a different type of problem, in which no baseline or regime information was provided [53]. In this case study, the previous developed signal processing and feature extraction techniques for gearbox condition monitoring are leveraged with the unique residual clustering health assessment technique.

6.1 Introduction

Condition monitoring and health assessment of rotating machinery using vibration signals has been an area of research for many years. Much of the research has focused on detecting incipient levels of component damage. Detecting the early stages of component damage can provide enough of a maintenance window to take appropriate action prior to the complete failure of the system [13]. Even a machine without any faults, will have a vibration signature that is based on the system dynamics and the forces acting on the system. However, degraded components such as bearings, gears, and shafts, have different vibration signatures when compared to the baseline signature that can be detected with the use of the proper signal processing methods. A vast amount of signal processing and feature extraction techniques has been developed for detecting bearing [118], gear [119], and shaft [13] vibration signatures. However, one of the main challenges in rotating machine condition monitoring is the potential for the machinery to operate in a multitude of different operating regimes with a different
vibration signature in each regime. In addition, the defect signatures could be less pronounced or observable under certain load and speed conditions, and much more pronounced and observable under other load and speed regimes [120]. This aspect of a signature that is only present in a subset of the operating conditions requires a unique health assessment methodology that incorporates this aspect. The residual clustering method that was previously developed for detecting a regime dependent anemometer sensor fault is applied to a similar regime based problem for gearbox health monitoring. A description of the data set is provided in Section 7.2, which is followed by an in depth study on the algorithmic processing steps applied to this data set in Section 7.3. A comparison between the results using the residual clustering method and the multiple regime density overlap method are provided in Section 7.4. Lastly, some concluding remarks and suggestions for further study are discussed in Section 7.5.

6.2 Data Description

The residual clustering approach was further applied to a condition monitoring application for a two stage gearbox. A schematic of the gearbox is provided in Figure 43, in which the system consists of four spur gears, 6 rolling element bearings, and three shafts. Two accelerometers are placed on the gearbox housing, with an accelerometer placed on the input side of the gearbox and also an accelerometer on the output side of the gearbox. A tachometer signal is also mounted on the input shaft of the gearbox and provides a 10 pulse per revolution pulse train for measuring the input shaft speed. A listing of the number of gear teeth for each of the respective gears is provided in Table 15. The speed ratio from the first stage of the gearbox is 3:1 and 1.67:1 for the second stage; thus the speed ratio between the input and output shaft is 5:1.
The gearbox was tested in 10 different discrete operating conditions that are listed in Table 16. The input shaft was run at 5 different rotational speeds (30Hz-50Hz) and also the torque load was adjusted from low to high [121]. In addition to these 10 operating conditions, the gearbox was tested under different condition monitoring cases, in which a case would represent a set of health states for the 6 bearings, 4 gears, and input and output shaft. A listing of the different condition monitoring cases is provided in Table 17 in which the health state of each component is also listed. As one can observe, the first case represents the baseline condition, in which all the bearing, shaft, and gear components are in a fault free (FF) health state. The other cases consist of single or multiple faults occurring on the different shaft, bearing, and gear components. There are 7 different seeded fault cases and a replication of each seeded fault
test was performed for a total of 14 runs. For each experimental run, 10 files were collected, providing a data file for each operating regime.

Table 16: Listing of Discrete Operating Regimes for Experimental Gearbox

<table>
<thead>
<tr>
<th>Regime #</th>
<th>Input Shaft Speed (Hz)</th>
<th>Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>Low</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>High</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>Low</td>
</tr>
<tr>
<td>7</td>
<td>45</td>
<td>High</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>Low</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>High</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>Low</td>
</tr>
</tbody>
</table>

Table 17: Listing of Seeded Faults for Each Experimental Test Case:

Highlighted in bold are examples used to demonstrate the monitoring approach

<table>
<thead>
<tr>
<th>Case #</th>
<th>Gear</th>
<th>Bearing</th>
<th>Shaft</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-11</td>
<td>FF</td>
<td>FF</td>
<td>FF</td>
</tr>
<tr>
<td>#2-3</td>
<td>Chipped</td>
<td>FF</td>
<td>Eccentric</td>
</tr>
<tr>
<td>#4-5</td>
<td>FF</td>
<td>FF</td>
<td>Eccentric</td>
</tr>
<tr>
<td>#6-7</td>
<td>FF</td>
<td>FF</td>
<td>Eccentric</td>
</tr>
<tr>
<td>#8-9</td>
<td>Chipped</td>
<td>FF</td>
<td>Eccentric</td>
</tr>
<tr>
<td>#10-11</td>
<td>FF</td>
<td>FF</td>
<td>FF</td>
</tr>
<tr>
<td>#12-13</td>
<td>FF</td>
<td>FF</td>
<td>FF</td>
</tr>
<tr>
<td>#14-15</td>
<td>FF</td>
<td>FF</td>
<td>FF</td>
</tr>
</tbody>
</table>
In order to illustrate the residual clustering monitoring approach, two examples from this set of seeded fault gearbox data are used. They consist of detecting an eccentric gear (Gear 3) and also an imbalance shaft (input shaft). For each example, a specific set of signal processing and feature extraction methods are used as inputs into the residual clustering approach. It should be noted that a specific feature set along with the same residual clustering approach could be applied to the other faulted gear, bearing, and shaft components. However, the intent of the case study was to demonstrate the monitoring approach for a mechanical system operating under discrete operating conditions and not to evaluate the method for detecting the entire set of degraded shaft, gear, and bearing components.

Each data file consisted of three columns of data (two accelerometers and tachometer signal) in which the signals were acquired at a sampling rate of 66.67 KHz and recorded for 4 seconds [122]. As a way of providing additional samples for training the residual processing algorithm, the signals were processed in two second overlapping analysis blocks that were incremented by 0.3 seconds. For a 4 second data file, this provided 6 processing blocks (6 samples), in which a feature vector was extracted from each processing block. Considering that there are 10 operating regimes and 10 data files for each health case, this provided 60 samples for each fault case. It should be noted that the baseline case had a replication run and consisted of 120 samples. However, 60 of them were randomly selected for training the residual processing algorithm and the remaining 60 were used for testing. In total, there are 960 samples in which 120 samples are from the baseline condition and half of which are used for training the model (900 total samples for testing). A single health value is processed for the 60 data samples for each experimental test case in Table 17. In this application, the monitoring period is selected to include data from all 10 operating regimes since the signature could be present in only a subset of the operating conditions.
6.3 Data Processing Methods

An overview of the data processing methods for monitoring the gearbox using the residual clustering approach is provided in Figure 44. The flow chart is quite similar to the previous anemometer case study, with a few key differences. Using the residual clustering methodology for gearbox monitoring puts more emphasis on the specific signal processing and feature extraction methods for detecting the different gear, shaft, and bearing failures. In this case study, the signal processing and feature extraction methods include performing time synchronous averaging (TSA) and narrowband demodulation for detecting the eccentric gear and shaft imbalance faults. The method also uses a multivariate figure of merit health value. The motivation for including a multivariate health metric is that fusing the information from the two accelerometers could offer better detecting of the gear and shaft problems as opposed to using a single feature from one accelerometer.

![Diagram of data processing methods](image)

Figure 44: Residual Clustering Health Monitoring Approach for Gearbox
The previous anemometer case study was focused on sensor fault detection and it was more applicable to have a univariate figure of merit health value for each sensor. The flow chart also includes generating a regime monitoring map after processing the health values. Effectively this allows one to know which set of operating regimes are most conducive for detecting the signature of a defective gear or shaft component.

6.3.1 Signal Processing and Feature Extraction

For each component and failure mode, a set of features are extracted using well established signal processing methods from the literature [13]. The signal processing method for detecting the eccentric gear uses synchronous averaging and amplitude modulation. For shaft imbalance, the signal processing methods include synchronous averaging and frequency domain analysis. Prior to presenting the details and example results from the signal processing methods for the eccentric gear example, a common signal model for gear vibration is reviewed. This provides some background information on the vibration signature of a normal gear and how local or distributed gear defects can change the meshing pattern and result in amplitude and phase modulation effects. Following the eccentric gear signal example, the signal processing methods for the shaft imbalance example are presented.

For a gear pair that is under constant load and speed conditions, the meshing vibration, \(x(t)\), consists of the fundamental gear mesh frequencies and several harmonics as shown in Equation 41. In this formulation, the number of teeth is denoted by \(G_T\), the shaft speed is denoted by \(f_s\), \(\phi_m\) is the initial phase, \(X_m\) is the amplitude of the \(m^{th}\) harmonic frequency, and \(M\) is the number of gear mesh frequency harmonics included [34].

\[
x(t) = \sum_{m=0}^{M+1} X_m \cos(2\pi m G_T f_s t + \phi_m)
\]  

(41)
The vibration signature changes for a distributed gear fault (eccentricity, wear on gear teeth) and also for local gear faults (chipped gear tooth, tooth root crack, spalls, pitting). These defects produce changes in the vibration signal and can be defined by their amplitude and phase modulation functions $a_m(t)$ and $b_m(t)$. Considering that the vibration is periodic with respect to the gear shaft rotation, the amplitude and phase modulation functions can be represented by a discrete Fourier series as indicated by Equation 42-43.

\[
a_m(t) = \sum_{n=0}^{N} A_{mn} \cos(2\pi n f_s t + \alpha_{mn})
\]

\[
b_m(t) = \sum_{n=0}^{N} B_{mn} \cos(2\pi n f_s t + \beta_{mn})
\]

Including the amplitude and phase modulation functions into the original expression provides a signal model as shown in Equation 44 when a gear defect is present.

\[
y(t) = \sum_{m=0}^{M+1} X_m \left(1 + a_m(t)\right) \cos\left(2\pi m G_f f_s t + \phi_m + b_m(t)\right)
\]

A narrow band pass filter is used to filter the signal around a selected gear mesh frequency or harmonic. With the assumption that the interference from neighboring gear mesh harmonic peaks is negligible, the filtered signal is given by the expression in Equation 45.

\[
z_m(t) = X_m \left(1 + a_m(t)\right) \cos\left(2\pi m G_f f_s t + \phi_m + b_m(t)\right)
\]

Further processing of the filtered signal can be considered by taking the Hilbert Transform and representing the result as an analytical signal in Equation 46.

\[
c_m = z_m(t) + jH(z_m(t))
\]
Considering that the Hilbert Transform of a cosine function is simply a sine function, and substituting the previous expression for \( z_m \) in Equation 45, \( c_m \) is given by Equation 47, and further simplified in Equation 48.

\[
c_m = X_m (1 + a_m(t)) \left[ \cos(2\pi m G_T f_s t + \phi_m + b_m(t)) + j \sin(2\pi m G_T f_s t + \phi_m + b_m(t)) \right]
\]  

(47)

\[
c_m = X_m (1 + a_m(t)) \exp\left(j \left(2\pi m G_T f_s t + \phi_m + b_m(t)\right)\right)
\]  

(48)

Estimations of the amplitude \( (a_m) \) and phase modulation \( (b_m) \) functions can be obtained by considering the modulus and phase of the analytical signal \( (c_m) \) and are provided in Equation 49 and Equation 50 respectively.

\[
a_m = \frac{|c_m|}{X_m} - 1
\]  

(49)

\[
b_m = \tan^{-1}\left(\frac{\text{imag}(c_m)}{\text{real}(c_m)}\right) - (2\pi m G_T f_s t + \phi_m)
\]  

(50)

As highlighted by the review of the gear vibration signal model, extracting the amplitude and phase modulation signals are imperative for detecting local or distributed gear defects. A flow chart that summarizes the processing steps for extracting the amplitude and phase modulation signals is outlined in Figure 45. This is a general flow chart for detecting distributed or local gear defects. For localized gear defects such as a spall, both the phase modulation and amplitude modulation signal are quite suitable for detecting that signature; with the phase modulation signal in certain instances providing a more incipient detecting than the amplitude modulation signal [123]. For distributed gear faults such as an eccentric gear, the amplitude modulation signal contains the signature information [124]. Considering this aspect, the example plots and calculated features are from the amplitude modulation signal.
Taking into account that Gear 3 is on the idler shaft, the vibration signals were resampled with respect to the idler shaft and averaged over several rotations during the two second analysis block. Example results from the time synchronous average (TSA) signal for the idler shaft are provided in Figure 46 for Case 1 and for Case 4. In Case 1, all the bearing, shaft and gear components are in a healthy state, while in Case 4, the only fault is an eccentric Gear 3. The top left and right graphs show the TSA signal for Case 1 and Case 4 at the lowest input shaft rotational speed of 30Hz and under high load. There are very minute differences in the signature for the two cases at this low speed, with Case 4 having a slightly larger peak to peak value. This is in sharp contrast to the signature at the highest shaft speed of 50Hz, in which Case 4 has a clear amplitude modulation signature and much larger vibration amplitude.

However, this vibration signature includes all vibration synchronous with the idler shaft and the narrowband amplitude modulation signal would extract features that are specific to Gear 3.
The amplitude modulation signal was obtained by using a band pass filter that included 3 sidebands around the gear mesh frequency (Orders 45-51) and then using the Hilbert Transform for obtaining the envelope of the filtered signal. The amplitude modulation signal for Case 1 and 4 is shown in Figure 47, in which it is quite difficult to discriminate between these two cases at a low speed (30Hz) but the signature is very clear at the highest operating speed (50Hz). This agrees with the original assumption that many mechanical signatures are regime dependent, in which a particular loading or speed condition might be more optimal for determining that one of the components is in a degraded condition.

Figure 46: Time Synchronous Average Signal Idler Shaft, Left - Case 1 (30Hz and 50Hz Shaft Speed, High Load), Right - Case 4 (30Hz and 50Hz Shaft Speed, High Load)
As an additional example of the eccentric gear signature, the vibration signal for Case 4 for the highest operating speed and load was plotted for 10 rotations in Figure 48. In this time signal, one can clearly observe the amplitude modulation that occurs for every rotation of the idler shaft. This inspection of the time waveform was used to confirm the previous signal processing results using synchronous averaging and narrowband demodulation. Although one can use the raw time waveform for feature extraction, the use of synchronous averaging would reduce the random noise and allow one to focus on vibration that is only synchronous with the idler shaft. The narrowband demodulation method allows one to further examine the vibration from an individual gear and thus the extracted features are representative of the health condition of only that specific gear.
The signal processing methods for detecting shaft imbalance involves fewer steps, and consists of frequency domain analysis of the TSA vibration signal [125]. In this example, the vibration signal was averaged with respect to the output shaft since the input shaft is a direct multiple of the output shaft (5:1) and would allow one to further extract frequency domain information for both shafts. Example frequency domain plots are provided in Figure 49 from a gearbox with a healthy input shaft (Case 1) and a gearbox with an imbalanced input shaft (Case 10). The top plots are when the input shaft is at 45Hz. At this particular shaft speed, the peak related to imbalance at 1X is actually lower in magnitude for Case 10 when compared to Case 1. However, the imbalance 1X peak is larger in magnitude for Case 10 when the shaft speed is
at the highest operating speed of 50Hz. This example highlights that detecting shaft imbalance is an easier task at the highest rotational speed. This agrees with the physical aspects of the shaft imbalance fault, in that the excitation force is proportional to the square of the shaft speed. However, this is just the excitation, and the response measured from the accelerometer would include effects from the transfer path and also vibration from other sources [126].

Figure 49: Time Synchronous Average Spectrum, Left - Case 1 (45Hz and 50Hz Shaft Speed, High Load), Right - Case 10 (45Hz and 50Hz Shaft Speed, High Load)
Table 18: Features for Eccentric Gear and Shaft Imbalance Health Monitoring

<table>
<thead>
<tr>
<th>Number</th>
<th>Signal</th>
<th>Signal Processing Method</th>
<th>Feature</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Accelerometer 1</td>
<td>TSA Idler Shaft- NB Envelope</td>
<td>Peak to Peak</td>
<td>Gear 3 - Eccentric</td>
</tr>
<tr>
<td>2</td>
<td>Accelerometer 2</td>
<td>TSA Idler Shaft- NB Envelope</td>
<td>Peak to Peak</td>
<td>Gear 3 - Eccentric</td>
</tr>
<tr>
<td>1</td>
<td>Accelerometer 1</td>
<td>TSA Output Shaft - FFT</td>
<td>Input Shaft 1X Peak</td>
<td>Input Shaft Imbalance</td>
</tr>
<tr>
<td>2</td>
<td>Accelerometer 2</td>
<td>TSA Output Shaft - FFT</td>
<td>Input Shaft 1X Peak</td>
<td>Input Shaft Imbalance</td>
</tr>
</tbody>
</table>

A tabular summary of the extracted features for the eccentric gear fault and the shaft imbalance fault are provided in Table 18. The peak to peak values from the amplitude modulation signal from both accelerometers represent the selected features for detecting the eccentric gear fault. The frequency domain peaks at 1X from both the input and output accelerometers are the selected features for detecting input shaft imbalance. In both cases, two features are used as inputs into the residual clustering health calculation. However, more or less features could be used for a given application depending on the type of failure mode and the number of available sensors. An example plot of the features from the baseline data set (Case 1) is provided in Figure 50 for the two eccentric gear features. In general the vibration features are higher in magnitude with higher operating speed. However, the magnitude of the features is actually the largest under a light loading condition and 50Hz input shaft speed and not with the higher torque load. Modeling the relationship among the features in a baseline state is done using the auto-associative neural network model which learns the correlation relationship between these vibration features. Other modeling options could include using a regression model to map the load and speed to the measured vibration features. However, this would not be a trivial task due to the potential nonlinear relationship between the vibration features and the operating regime variables [127].
The input shaft imbalance features for the different operating regimes from Case 1 are shown in Figure 51. The features have the largest magnitude under the highest operating speed and the high load regime. However, the shaft vibration features are not the smallest in the 30Hz and low torque load regime. For example, the input shaft 1X Peak has the smallest magnitude for the first accelerometer in the 35Hz and high load operating regime. Based on the inspection of the shaft vibration features, modeling the correlation relationship using a nonlinear model would be an appropriate choice [92]. In this case study, the nominal relationship between the shaft vibration features is modeled using an auto-associative neural network.
6.3.2 Residual Processing and Figure of Merit Results

The trained auto-associative neural network model is used to calculate predicted values for the amplitude modulation peak to peak values. Comparison of the predicted and actual feature values are provided for Case 1, Case 12, and Case 4 in Figure 52. There are 60 test samples for each health case, in which there are 6 samples from each of the 10 operating regimes. The samples are arranged from the lowest to highest operational speed. The first 6 samples from each shaft speed are from the high load setting and the remaining 6 samples are from the low load regime. The results for Case 1 show a very close match between the predicted and actual feature values for all 10 operating regimes. This should be expected, since Case 1 represents the baseline condition and one would expect the model to provide accurate predictions when the gearbox system is in a healthy state. The middle plot in Figure 52 is from Case 12, in which Gear 3 is healthy but there is a damaged bearing component and the output shaft has a faulty
keyway. However, the amplitude modulation features are specific to Gear 3. This agrees with the residual processing results, in which the predicted and actual feature values are in close agreement for Case 12. The results for Case 4, in which the only fault for the gearbox system is an eccentric Gear 3, are quite encouraging. The predicted and actual feature values are a close match for the first 18 samples, which are from data collected at a shaft speed of 30Hz and 35Hz respectively.

Figure 52: Predicted and Actual AM Peak to Peak Gear 3 Features from Accelerometer 2, Top Plot – Case 1, Middle Plot – Case 12, Bottom Plot – Case 4 (Eccentric Gear)

There are some slight deviations between the predicted and actual values for the medium operating regime data. The residual values are quite large in magnitude for the last 12 samples which are when the input shaft has a shaft speed of 50Hz. The feature values are larger than what the model predicts for the highest operating speed, which agrees with the notion that a fault would cause larger amplitude modulation and vibration than expected. The results again
agree with the hypothesis that many mechanical faults signatures are regime dependent and are more noticeable under certain working conditions.

The data is further processed by clustering the residual values into two clusters and selecting the cluster with the more positive residual summation value. Selecting the cluster with the more positive residuals is based on the notion that vibration features that are larger than normal are representative of a fault, while vibration features lower than predicted does not signify a problem but rather normal variation [89]. The figure of merit health metric is the calculated residual summation value from the selected cluster, in which a larger value would imply a worse Gear 3 health condition.

Three different clustering algorithms were evaluated in this study: k-means, a density based Gaussian mixture model (GMM) algorithm [128], and a hierarchical clustering algorithm [129]. Hierarchical clustering is based on the dissimilarity in the features and allows one to visual examine the cluster structure. In this case study, the hierarchical clustering algorithm used a Euclidean based similarity matrix and a centroid based linkage distance calculation. An example result using hierarchical clustering to cluster the residual feature is presented in Figure 53, in which the example is from Case 4, which had an eccentric Gear 3 fault. From this diagram one can observe that there are two clear clusters in the residual data features and a cut-off value of approximately 0.7 to 1.2 would result in two clusters. The diagram also shows that three clusters could be formed if one chooses a cut-off value of approximately 0.5 to 0.75; however it appears from the diagram that two clusters is a more appropriate choice. The assumption with the residual clustering approach is based on the idea that there would be two clusters in the residual features, since there would be a regime in which the signature is more pronounced and an operating regime in which the signature is less distinguishable. Considering this assumption, for each of the three clustering algorithms, the algorithm was configured to find 2 clusters in the residual features.
For k-means and GMM, five replications are used during the cluster estimation process in order to avoid a local minimum being selected. The figure of merit health values using k-means, GMM, and hierarchical clustering are provided in Figure 54 – Figure 56. The results using any of the three clustering algorithms show that the residual clustering method provides a clear way to discriminate between an eccentric gear and a gear in a healthy state. The figure of merit values for the cases in which Gear 3 is in a healthy condition are all quite low and at a normal level. It should be noted that the features used in this example are solely designed for detecting an eccentric Gear 3. This goes along with the figure of merit results, in which the values are quite low for cases in which other gear, bearing and shaft components are damaged, but Gear 3 is in a healthy state. The method can detect all 8 cases in which Gear 3 is in a faulty eccentric condition. The figure of merit values are the largest in magnitude when the eccentric Gear 3 is meshing with a damaged Gear 4 in Cases 6-9. This implies that the degradation signature of the eccentric gear could be amplified when it is meshing with a gear that also has a fault.
Figure 54: Figure of Merit Health Value for Detecting Eccentric Gear 3 Using Auto-Associative Neural Network (AANN) and K-means clustering

Figure 55: Figure of Merit Health Value for Detecting Eccentric Gear 3 Using Auto-Associative Neural Network (AANN) and Gaussian Mixture Model Clustering
The differences in the health values when k-means clustering, GMM, or hierarchical clustering is used are very minute; all three methods had good detection results for the eccentric gear fault. The results from this example show that the selection of the clustering algorithm did not have a significant effect on the monitoring results. A more detailed comparison of the health monitoring results using the three different clustering algorithms is discussed in Section 6.4.

In addition to considering different clustering methods for the residual clustering approach, it is important to consider how the approach compares to other multi-regime health monitoring methods. A recently developed, distribution based multiple regime health monitoring method that has been applied to wind turbine performance monitoring [78] and also for bearing
components in a chiller compressor [77] is used to benchmark the residual clustering approach. The distribution method is based on comparing the overlap in the distribution between a baseline state and the current state, in which each health state can be represented by a mixture of Gaussians. Modeling the baseline feature distribution, \(H(x)\), by a GMM model is given by the expression provided in Equation 51, in which \(R\) is the number of mixtures, \(p_i\) is the weight for the \(i^{th}\) mixture, and \(\mu_i\) and \(\sigma_i\) are the mean and standard deviation for the \(i^{th}\) mixture. The feature distribution of the monitored system \(G(x)\) is also modeled in a similar manner to \(H(x)\), in which the same number of mixtures is used to describe the feature distribution.

\[
H(x) = \sum_{i=1}^{R} p_i N(\mu_i, \sigma_i)
\]  

(51)

The comparison between the two feature distributions is computed using the \(L_2\) distance, and further normalized by the \(L_2\) distance value for each distribution as shown in Equation 52. The interested reader is referred to the dissertation by Linxia Liao [130], in which a more detailed description of the calculation procedure is provided in the appendix of his dissertation.

\[
CV = 1 - \frac{\|H(x) - G(x)\|_{L_2}}{\sqrt{\int \left( (H(x))^2 + (G(x))^2 \right) dx}}
\]  

(52)

The overlap value (CV) is bounded between 0 and 1, with a value of 1 indicating complete overlap (healthy) and a value closer to 0 indicating very little if any overlap in the distributions. To agree with the convention used by the residual clustering figure of merit health value in which a higher value implies worse health, a health metric for the overlap method is defined in Equation 53 in which one minus the CV value is used.

\[
h_i = 1 - CV
\]  

(53)
For implementing the GMM overlap method, the distribution of the residuals in the baseline state is used to compare the residual distribution for each test case. The results using this method are provided in Figure 57, in which the health values for the eccentric gear cases (2-9) are indicated in green and the cases in which Gear 3 is healthy (1, 10-16) are indicated in blue. Although there is some separation in the health values between a healthy and eccentric Gear 3, there is much more variation in the health values for the cases in which Gear 3 does not have a fault. A more formal comparison between the residual clustering approach and the GMM overlap method is provided in Section 7.4. However, the initial observations indicate that the GMM overlap health value has more variation and less separation between the eccentric gear and healthy gear state. One potential drawback of the GMM overlap method is that it compares the residuals in all 10 operating regimes; it does not effectively consider that the signature is
present in only a subset of the operating regimes. In addition, it does not consider that positive residuals (vibration features larger than normal) are more representative of a degraded condition and instead weights both positive and negative residuals equally in the overlap calculation.

The residual processing results for the shaft imbalance feature set are provided in Figure 58. The predicted and measured shaft imbalance features show a close match when the input shaft is in a fault free condition for Case 1 and Case 2. This is quite different then the results for Case 10 when there was an imbalanced input shaft. The bottom plot in Figure 58 shows that the measured shaft imbalance features are larger than the predicted feature values in the highest speed regime (50Hz). The results highlight that the shaft imbalance signature is dependent on the operating conditions and is much easier to detect at a higher shaft speed.

Figure 58: Predicted and Measured Input Shaft 1X Vibration Features from Accelerometer 1, Top Plot – Case 1, Middle Plot – Case 2, Bottom Plot – Case 10 (Eccentric Gear)
Following the same clustering and figure of merit health calculation procedure that was used for the eccentric gear example, the figure of merit health values were calculated for the shaft imbalance fault. The figure of merit input shaft health values using k-means, GMM clustering, and hierarchical clustering are provided in Figure 59 – Figure 61. The residual clustering figure of merit values can clearly discriminate between a gearbox with a healthy input shaft and a gearbox that has imbalance in the input shaft. The health values for Cases 10-11 and Cases 14-15 are clearly higher than the values when the input shaft is fault free. The results using the three different clustering algorithms are quite similar; this implies that the selection of the clustering method did not have a significant effect on the output health value.

Figure 59: Figure of Merit Health Value for Detecting Input Shaft Imbalance Using Auto-Associative Neural Network (AANN) and K-means clustering
Figure 60: Figure of Merit Health Value for Detecting Input Shaft Imbalance Using Auto-Associative Neural Network (AANN) and Gaussian Mixture Model Clustering

Figure 61: Figure of Merit Health Value for Detecting Input Shaft Imbalance Using Auto-Associative Neural Network (AANN) and Hierarchical Clustering
Although the figure of merit results for Cases 12-13 are lower in magnitude than the cases where the input shaft has imbalance, they are higher than the other input shaft fault free health values. One possible explanation is that Cases 12-13 have a bad keyway on the output shaft. This would cause more fluctuation in the input shaft speed and effect the synchronous averaging and the extraction of the frequency domain peaks. This however is a minor concern, in that the figure of merit health values for cases 12-13 are still 3 to 4 times lower in magnitude than the health values when the input shaft has imbalance.

![GMM Overlap CV Value – Shaft Imbalance](image)

*Figure 62: Distribution Overlap Eccentric Gear 3 Health Value Using Gaussian Mixture Model*
The residual clustering method was compared with the multiple regime GMM overlap method for the shaft imbalance case study. The results using the GMM overlap method are provided in Figure 62, in which there is minimal separation between the healthy gearbox system and one with shaft imbalance. This highlights some of the potential disadvantages of the GMM overlap method, in that calculating the overlap in the distributions for the entire operating space is not an effective method for many common faults that can occur in mechanical systems. This is because the degradation signatures are very difficult to detect in certain operating regimes. In this shaft imbalance example, the signature is much easier to detect at higher operating speeds such as 50Hz than the medium and lower shaft speeds. The residual clustering method takes into account the aspect of a regime dependent signature, while the GMM overlap method does not effectively consider this issue.

6.4 Comparison of Methods and Regime Monitoring Map

The initial observations suggested that the residual clustering method was a more effective algorithm for this gearbox case study. There are various ways of comparing algorithm detection and classification performance including calculating receiver operating curves (ROC), fisher values, and confusion matrices [131]. For this case study, the detection results were compared using the fisher value, which would provide a measure of separation between the normal and faulty data sets. The fisher value can be calculated using the expression in Equation 54, where P and Q are the two health classes (normal and faulty), \( \mu_{P,i} \) and \( \mu_{Q,i} \) are the means of the \( i \)th health metric for each class, and \( \sigma^2_{P,i}, \sigma^2_{Q,i} \) are the variances [132].

\[
J_f(P, Q) = \frac{\left| \mu_{P,i} - \mu_{Q,i} \right|^2}{\sigma^2_{P,i} + \sigma^2_{Q,i}} \tag{54}
\]
The results of applying the fisher criterion metric to the eccentric gear example are indicated in Figure 63, in which one can observe that the GMM overlap had the smallest fisher value. This indicates that the GMM overlap method had the worst performance and smallest separation between the two health classes. The residual clustering method offered superior detection results with a much larger fisher value. Regarding the residual clustering method, similar performance was achieved using the k-means and GMM clustering algorithm and slightly worse performance was achieved when hierarchical clustering was used. It appears that one aspect related to the improved performance of the residual clustering approach can be attributed to the fact that the health metric is calculated from a subset of the operating conditions. The GMM overlap method is calculated using data from the entire operating regime.
space and this can have a detrimental effect on the detection performance if the signature is not present for all the operating regimes.

The number of files (D), in which the fault are detected can be used for formulating a regime monitoring map that highlights which operating regime is best for detecting the degradation signature. The regime map calculation provided in Equation 55, considers the number of samples in each operating regime (G\text{ij}) in the selected cluster. The calculation result is further normalized by the number of files when a fault was detected (D) and the number of analysis blocks in each file (L\text{s}). The number of samples in operating regime from the selected cluster is provided in Table 19 for the eccentric gear case study. In this example, 8 files (Cases 2-9) were detected as having a fault, and each file was divided into 6 analysis blocks; thus the maximum number of samples in each regime is 48. The normalized values can be visualized in a regime map, indicating the percentage of samples in that regime in which the signature is present.

\[
\text{Map}(i) = \frac{\sum_{j=1}^{D} G_{ij}}{D \cdot L_s} \quad \text{for } i = 1 \text{ to } 10
\]

(55)

Table 19: Number of Samples in Each Regime in Selected Cluster for the Eccentric Gear Fault

<table>
<thead>
<tr>
<th></th>
<th>30Hz Speed</th>
<th>35Hz Speed</th>
<th>40Hz Speed</th>
<th>45Hz Speed</th>
<th>50Hz Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low Torque Load</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td><strong>High Torque Load</strong></td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>24</td>
<td>46</td>
</tr>
</tbody>
</table>
The regime monitoring map for the eccentric gear fault is presented in Figure 64, in which one can observe that the best monitoring regimes consists of the 45Hz and 50Hz shaft speeds. An interesting observation is that the regime map is indicating that the signature is more prevalent under light load instead of high torque load, when the shaft rotational speed is 45Hz. The monitoring map agrees with the previous feature extraction and residual processing results, in which the eccentric gear fault signature was not observed for the lower shaft speeds of 30Hz and 35Hz. This type of information can aid future design of prognostic and health management systems, in which local health models can be developed for only the best monitoring regimes.

The health monitoring results using the residual clustering approach and the GMM overlap method for the shaft imbalance fault were also evaluated using the fisher criterion. The fisher values are provided in Figure 65 and highlight the effectiveness of the residual clustering method. The residual clustering method provided a clear ability to discriminate between a fault
free shaft and a shaft with imbalance. The results using the residual clustering approach were quite similar, irrespective of whether k-means, GMM clustering, or hierarchical clustering was used. Considering that all three clustering methods provided similar results, it appears that the selection of the clustering algorithm did not have much influence on the output health values. The GMM overlap method provided inadequate results with a very low fisher value. This highlights that the GMM overlap method was not able to discriminate between the two shaft conditions, which is in sharp contrast to the residual clustering method which could clearly detect the shaft imbalance fault.

Figure 65: Fisher Value between Healthy and Fault Class - Shaft Imbalance

In order to further interpret the residual clustering method results, a regime monitoring map was calculated for the shaft imbalance fault. The regime monitoring map is provided in Figure
66, in which it clearly shows that the best operating regime is at a shaft speed of 50Hz and under either torque load. The regime monitoring map reinforces the previous results, where even at 45Hz shaft speed; the shaft imbalance fault could not be clearly detected. Unlike the eccentric gear example, the monitoring map for the shaft imbalance fault was more narrow and suggested only two of the ten operating regimes were conducive for detecting this type of fault. Considering that the signature is only present in 20% of the data samples, comparing the distributing for the entire operating space might not provide an adequate detection. This agrees with the superior detection results obtained by the residual clustering method when compared with the GMM overlap method.

Figure 66: Regime Monitoring Map for Shaft Imbalance Fault
6.5 Discussion of Results

The residual clustering approach was further applied to the gearbox case study with very encouraging detection results. The approach consisted of extracting specific gear and shaft features using established signal processing and feature extraction methods, calculating predicted feature values from baseline models, clustering the residual values, and calculating a figure of merit health value from the selected cluster. The residual clustering method provided more discrimination for both the shaft imbalance and eccentric gear fault when compared with the GMM overlap method. This case study reaffirms the original hypothesis that improved monitoring results are obtained by clustering and focusing on a particular subset of the operating regimes in which the signature is most prevalent. In addition, the gearbox case study illustrated how the residual clustering results can be used to calculate a monitoring map. The regime monitoring map for the eccentric gear fault indicated that operating speeds of 45Hz and 50Hz and a light torque load was conducive for detecting that type of gear problem. The shaft imbalance fault was best detected under the highest shaft rotational speed of 50Hz. Further study could consider the application of the residual clustering approach to mechanical systems with continuous and time varying operating conditions.
CHAPTER 7: COMPONENT DEGRADATION EXAMPLE – WIND TURBINE DRIVETRAIN CONDITION MONITORING CASE STUDY

The third and final case study is used to benchmark the residual clustering methodology with other existing approaches for assessing the health condition of a multi-regime system. It also is used to demonstrate how the residual clustering approach can be used to monitor the degradation of an asset over time. This case study is for an offshore wind turbine, in which vibration and SCADA data is provided for a 15 month time period. In addition, the wind turbine develops a problem in the rotor, in which downtime occurs and the turbine is not operational for a period of time. The residual clustering method is challenged with not only detecting the rotor issue, but also providing an early warning and trend of the potential problem.

7.1 Introduction

Despite year to year variations, there has been considerable growth in the number of installations of wind turbines. Based on 2012 numbers, the total number of installed capacity for wind turbines reached 282.5 GW, and there were a record number of installations in both Europe and North America [133]. Although the growth in installation capacity is quite encouraging, improvement in operating and maintenance practices are needed for making wind energy more competitive when compared with other fossil fuel power generation options. For reducing operating and maintenance cost, the use of a condition monitoring system is one option that could offer this desired reduction in maintenance cost. For offshore wind turbines, the need for a condition monitoring system is even more prevalent. The cost associated with an offshore wind turbine gearbox failure is greater than the gearbox failure cost for onshore turbines, since there is greater difficulty in performing the maintenance out in sea. Even though the offshore wind turbine market is non-existent in North America, there was 1,166MW of
installed offshore wind capacity in Europe [134], and the potential growth of the worldwide offshore wind turbine market places even more importance on the need to have robust analysis methods for condition monitoring systems.

Considering the importance of condition monitoring for wind turbine applications, there has been considerable prior research and development work in this area. A discussion on the appropriate hardware and software architecture for a wind turbine condition monitoring system was presented by Hameed et al. [135]. A comprehensive report was also provided by the National Renewable Energy Lab (NREL), in which several university and industry partners contributed and highlighted there vibration based monitoring methods and algorithms for wind turbine gearbox components [136, NREL report]. Recently, there also have been specific studies focused on offshore wind turbine condition monitoring applications. A study performed by Miguelanez et al. [137] presented a holistic approach for offshore wind turbine monitoring, in which outputs and signals from different parts of the turbine are fused together for providing the component level diagnosis. A study by Zhixin and Yue [138] focused on the use of a wireless sensor architecture for monitoring offshore wind turbines. The rationale for considering the wireless network was based on how far from land the offshore wind turbines are and the difficulty in setting up the communication infrastructure for a traditional wired architecture. In addition, a recent study by Zhao et al. [139] used advanced vibration feature extraction methods and a health indicator based on a self-organizing map to monitor the health condition of key rotating components in an offshore wind turbine. The results from the recent study by Zhao [139] highlight that data driven condition monitoring methods are applicable for offshore wind turbine applications, but considerable work is needed to understand whether other multi-regime health monitoring algorithms could have achieved similar or better performance. The work in this dissertation investigates the residual clustering health monitoring approach and three other multi-regime health monitoring methods for an offshore wind turbine application; Sections 7.2 –
Section 7.5 include a description of the problem, the analysis methods, and the results from each algorithm.

7.2 Data Description

The objective of this study is to develop analysis methods for monitoring a 3 MW offshore wind turbine. The focus of the monitoring system is on the drivetrain system, in which it consists of three stages. The first and second stage consists of a sun gear, planetary gears, and an output sun pinion; the third stage is a parallel stage gearbox with two gears. A schematic of the drivetrain, along with the location of the accelerometers, is provided in Figure 67. The input in the drivetrain is the rotor, in which typical input speeds of the rotor are approximately 7 to 16 rotations per minute. The generator shaft is the output of the drivetrain system and there is a speed multiplication of 76.64 between the rotor and generator; the rotational speed of the generator is between 500 to 1200 rotations per minute.

![Gearbox Configuration and Sensor Location Diagram](image)

Figure 67: Gearbox Configuration and Sensor Location Diagram

Regarding the condition monitoring system, there are eight accelerometers installed on the drivetrain, with two accelerometers on the main bearing, four on the three stages of the gearbox, and two on the generator. A sampling rate of 6,250Hz is used for all eight
accelerometers, and the duration of one file is approximately 85 seconds. The condition monitoring system is configured to collect the vibration data once per day at midnight; however there are a few exceptions in which data is collected at a different time of the day or data is actually collected more than once per day. The wind turbine is also monitored by a supervisory control and data acquisition (SCADA) system, which monitors low frequency parameters such as power output, wind speed, rotational speeds of various shafts, and temperatures at various locations. The SCADA data is processed every 10 minutes, in which statistics such as the mean, standard deviation, maximum, and minimum are calculated during that 10 minute period. The data for this study was collected for a period of 15 months. During this period, a problem with the rotor shaft caused a downtime event that lasted for 2 weeks. The proposed health monitoring methods are evaluated with respect to accurately monitoring and detecting the rotor shaft problem before the maintenance event occurs.

7.3 Data Processing Methods

The first step in interpreting the raw data is to apply established signal processing and feature extraction methods to both the vibration data and the SCADA data. This includes the use of kurtosis based filtering methods, envelope analysis, extracting statistical parameters, and also the use of vibration data validation checks to remove erroneous data recordings. Section 7.3.1 describes these various signal processing and feature extraction methods that are applied to this offshore wind turbine application. Section 7.3.2 uses the proposed residual clustering health monitoring algorithm to assess the health condition of the wind turbine rotor shaft. Three other health assessment methods are applied to the same data set, in which the results are provided in Section 7.3.3 and a comparison of all four methods is provided in Section 7.4. Lastly, some conclusions from this wind turbine condition monitoring case study are provided in Section 7.5.
7.3.1 Signal Processing and Feature Extraction

Additional processing of the measured vibration signals can allow for more incipient detection of potential drivetrain problems and also more accurate and robust estimation of the health condition of the various components. For early detection, the use of filtering methods can be quite important, since the component degradation signature could be masked by the other vibration components and noise. A popular filtering method for early detection is based on a spectral kurtosis calculation, in which a kurtosis value is calculated as a function of frequency as shown in Equation 56. The filter is constructed using the kurtosis values for each frequency bin as shown in Equation 57; a given frequency bin is only included if the kurtosis value is above a statistical threshold at a set confidence level [18]. For applying the filter to the measured signal, a frequency domain multiplication is used, and the result is then transformed back to the time domain as indicted in Equation 58. The interested reader is referred to the work by Antoni et al. [18] for a more detailed discussion on the use of spectral kurtosis for machine condition monitoring.

\[ K_r(f) = \frac{\langle H^4(t, f) \rangle}{\langle H^2(t, f) \rangle^2} - 2 \]  
\[ \hat{W}(f) = \begin{cases} \sqrt{K_r(f)} & \text{for } K_r(f) > s_o \\ 0 & \text{Otherwise} \end{cases} \]  
\[ y(t) = \mathcal{F}^{-1} \left\{ \hat{W}(f)X(f) \right\} \]  

In this wind turbine drivetrain application, the spectral kurtosis filtering was used with a blocksize of 256 points. Example time domain vibration waveforms and the filtered vibration signals are provided in Figure 68 and Figure 69 for the axial main bearing and radial main bearing accelerometer. These example waveforms are 63 days before the maintenance event,
yet one can still observe a clear repetitive set of impacts for the axial main bearing vibration signal. These series of impacts are also seen in the raw signal for the axial bearing accelerometer. Observing these types of repetitive impacts in the raw radial main bearing vibration signal is quite difficult since the other vibration sources mask this problem quite well. However, the filtered signal shown in the bottom plot of Figure 69 clearly shows the similar repetitive impact signature that was observed in the axial main bearing accelerometer. The vibration waveform for the accelerometer near the first stage of the gearbox is provided in Figure 70; however it is too far removed from the fault location to pick up this problem at this particular time. These waveform plots highlight that the filtering method can be used to provide a clear visual observation of the vibration fault signature, and that the problem is likely the closest to the axial main bearing accelerometer.

Figure 68: Axial Bearing Vibration Signal - File # 62 (March 22, 2011)
Further visual observations of the vibration waveform were made from a time instance that was only 2 days before the maintenance event occurred. The raw and filtered vibration
waveform from the main bearing radial accelerometer is provided in Figure 71, and one can notice an even more pronounced set of repetitive impacts of in the vibration signal. It is quite interesting to note that the impacts can now be observed in the raw signal which is in contrast to the earlier recording shown in Figure 69.

![Waveform](image1)

![Filtered Signal](image2)

An indication of the severity of the problem can be noticed in the filtered vibration signal for the radial accelerometer located on the first stage gearbox. The filtered vibration waveform is shown in Figure 72 and one can observe the same repetitive impact signature; however it is quite difficult to observe this signature in the raw vibration signal. One can at least qualitatively conclude that the problem is more severe considering that the source of the problem is on the main bearing shaft but an accelerometer on the first gearbox stage is now picking up the signature.
Figure 72: Radial Stage 1 Gearbox Vibration – File #115 (May 14, 2011)

Figure 73: Envelope Vibration Spectrum - File #115 (May 14, 2011)
Additional qualitative analysis was conducted by examining the envelope spectrum of the vibration signal from the main bearing radial accelerometer. A plot of the envelope spectrum is shown in Figure 73 and one can observe significant peaks at 0.1192Hz and harmonics of that frequency. This frequency corresponds to 8.4 rotations per minute which corresponds to the rotor shaft speed. The envelope spectrum provides an additional set of evidence that the source of the vibration impact signature is from the wind turbine rotor shaft.

Although visually examining the vibration signals can be used to assess the condition of various components, they are not suited for an automated condition monitoring system since they require human expertise and interpretation. Statistical parameters from the raw vibration waveform, from the filtered signal, and from the envelope spectrum were extracted in order to be further processed by an automated health assessment wind turbine monitoring system. In addition, a vibration validation metric, based on the number of times a consecutive sample in a waveform is the same was calculated [3]. This was done in order to remove erroneous recordings in which the vibration signal was recorded when the wind speed or power setting was very low or zero. The SCADA recordings that were recorded at the nearest time instant to the condition monitoring system vibration recording were also stored, so one could know what was the rotational speed or power output for a given vibration acquisition.

An example plot of a vibration feature (axial vibration RMS), along with the vibration validation metric, and the generator speed are shown in Figure 74. One can observe several instances of higher than normal vibration level for the axial bearing accelerometer prior to the maintenance event which occurs on 5/16/2011. In addition, the vibration validation check shows that there are numerous instances of erroneous data files. Using a threshold of 480 for the validation check, all the files above that threshold are highlighted in blue. It is not surprising that the files that do not pass the vibration validation check also have unreasonably low vibration levels since the turbine is probably not running and the RMS magnitude is simply due
to sensor noise. It is quite interesting to notice on the bottom plot in Figure 74 that there are several instances in which the erroneous data files do not have a corresponding generator rotational speed of zero. Considering that the SCADA data is an average measurement for a 10 minute period, perhaps the generator was only idle for a portion of that 10 minute period in which the CMS data was also recorded. Regardless, it shows that even with speed and power information from the SCADA system, it is quite helpful to still have a signal validation check performed on the vibration signal to correctly include and exclude data files.

Figure 74: Feature Plot of Axial Vibration (Top Plot), Vibration Validation Metric (Middle Plot), and Generator Speed (Bottom Plot)
After removing the erroneous data files when the drivetrain system was idle, the calculated condition monitoring features could be further processed. An example plot of several condition monitoring features plotted over time is shown in Figure 75. The upper left plot in Figure 75 is the axial vibration peak to peak feature, in which one can observe some very abnormally high values right before the maintenance event. Despite these large values, there are only 5 recordings that are at this abnormally high value and thus it would be hard to use only this single feature to provide an earlier warning of the rotor shaft problem.

Figure 75: Monitored Parameters over Time, Axial Vibration Peak to Peak (Top Left), Radial Vibration Kurtosis (Bottom Left), Rotor Temperature (Top Right), Active Power (Bottom Right)
In addition, the rotor temperature shows a rather higher temperature prior to the maintenance event, but there are only 4 or 5 recordings prior to the maintenance event and thus not much of an advanced notice. The radial kurtosis feature trend, shown in the bottom left plot in Figure 75, provides a much earlier indication of the rotor problem. There are several high kurtosis values before the maintenance event. However, the kurtosis value has quite a bit of variation over time and the trend is not very monotonic.

The health assessment algorithms would look to fuse these condition monitoring features so that the attributes such as the kurtosis feature can help provide the incipient detection while other features could provide a more consistent trend over time. In addition, if one observes the plot of the output power, which is shown in the bottom right in Figure 75, a multi-regime health monitoring algorithm is needed since the operating variables have considerable variation. Four different health assessment algorithms are considered, based on their applicability to handle the variation in the operating variables and also for fusing the condition monitoring features into a single health value; specifics and results from these methods are presented in Section 7.3.2 and Section 7.3.3 respectively.

7.3.2 Residual Clustering Method and Results

The initial results from the signal processing and feature extraction methods showed that the filtering method had the potential to provide an earlier indication of the rotor shaft problem. Considering that aspect, a feature set was selected that included three statistics from the filtered axial bearing accelerometer and three statistics from the filtered radial bearing accelerometer. Also, since the residual clustering method is based on the prior assumption that the features are correlated, a correlation matrix between the features was calculated. The correlation matrix is provided in Table 20 for the six selected features and one can observe a high level of correlation between many of the features. Examples of high correlations include the axial and radial accelerometer RMS (0.74) and also the axial and radial accelerometer kurtosis (0.75).
Table 20: Correlation Matrix for Monitoring Features for Wind Turbine Rotor Condition

<table>
<thead>
<tr>
<th></th>
<th>Axial Main Bearing Vibration Peak to Peak</th>
<th>Axial Main Bearing Vibration RMS</th>
<th>Axial Main Bearing Vibration Kurtosis</th>
<th>Radial Main Bearing Vibration Peak to Peak</th>
<th>Radial Main Bearing Vibration RMS</th>
<th>Radial Main Bearing Vibration Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axial Main Bearing Vibration Peak to Peak</td>
<td>1.00</td>
<td>0.94</td>
<td>0.21</td>
<td>0.78</td>
<td>0.71</td>
<td>0.45</td>
</tr>
<tr>
<td>Axial Main Bearing Vibration RMS</td>
<td>0.94</td>
<td>1.00</td>
<td>0.02</td>
<td>0.75</td>
<td>0.74</td>
<td>0.35</td>
</tr>
<tr>
<td>Axial Main Bearing Vibration Kurtosis</td>
<td>0.21</td>
<td>0.02</td>
<td>1.00</td>
<td>0.15</td>
<td>-0.12</td>
<td>0.75</td>
</tr>
<tr>
<td>Radial Main Bearing Vibration Peak to Peak</td>
<td>0.78</td>
<td>0.75</td>
<td>0.15</td>
<td>1.00</td>
<td>0.86</td>
<td>0.47</td>
</tr>
<tr>
<td>Radial Main Bearing Vibration RMS</td>
<td>0.71</td>
<td>0.74</td>
<td>-0.12</td>
<td>0.86</td>
<td>1.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Radial Main Bearing Vibration Kurtosis</td>
<td>0.45</td>
<td>0.35</td>
<td>0.75</td>
<td>0.47</td>
<td>0.12</td>
<td>1.00</td>
</tr>
</tbody>
</table>

For applying the residual clustering method, an auto-associative neural network was trained using baseline data and the six calculated vibration features. The auto-associative neural network structure used 2 nodes in the bottleneck layer and four nodes in the mapping and demapping layer. The first 44 files (January 24 – March 11) were considered as baseline and used to train the auto-associative neural network algorithm so it could learn the nominal correlation relationship in the data. After the maintenance event, the algorithm was retrained.
since a major maintenance replacement usually results in a shift in the calculated features and a new baseline condition. In addition, the residual clustering algorithm is based on having sufficient samples for performing the clustering aspect. In most instances, the condition monitoring system would record several files in a single day, as opposed to a single acquisition per day that was used in this application. Considering the limited number of files, a moving window approach was used, in which a window of 18 files was used with an overlap of 16. This same moving window approach was also used for the other health assessment algorithms (GMM, SOM-MQE, PCA $T^2$), in which the distance based metrics would take an average value for the moving window samples.

After clustering, the cluster selection was based on which cluster had the greater residual mean summation value, since vibrations higher than expected would be more indicative of a problem as opposed to vibration features lower than predicted. The residual mean summation value using the six features was designated as the figure of merit health value and was calculated for each file. In addition to having a calculated health value, it is also necessary to have a corresponding threshold for alerting operators on when a problem is occurring that requires attention. Considering that the health value involves calculating residuals based on a non-linear neural network model and clustering, it is difficult to determine the underlying distribution of the health value. A more appropriate way of handling this aspect for threshold setting is to use Chebyshev’s inequality since that relationship is appropriate for any distribution. For an unknown distribution, Chebyshev’s inequality is given by the following expression provided in Equation 59, in which FOM is the figure of merit health value, $\mu$ is the mean of the health value, $\sigma$ is the standard deviation of the health value, and $k$ is a parameter that is directly related to the false alarm rate [140]. Effectively, the false alarm rate is given by $1/k^2$, in which a large value of $k$ would mean a lower false alarm rate and a threshold that is more standard deviations away from the mean. In this study, a $k$ value of 4 was selected which corresponds to
a false alarm rate of 0.0625, and a threshold that is 4 standard deviations greater than the mean health value. In addition, the health value mean and standard deviation are based on the first 20 calculations of the health value.

\[ \Pr(|FOM - \mu| \geq k\sigma) \leq \frac{1}{k^2} \]  

(59)

Figure 76: Residual Clustering Figure of Merit Health Value for Wind Turbine Rotor

It should be noted that the figure of merit health value was normalized by the threshold, so any values above 1 would be considered above the threshold and require further attention. The results of the residual clustering method are shown in Figure 76, in which one can notice the clear increasing trend in the health value until the maintenance event occurs. The health value crosses the threshold 26 days before the maintenance event occurs, so there would be almost 4 weeks of advanced notice of the problem. In addition, the health monitoring method provided
an estimate of the severity of the problem, as the health value increased until it was more than 20 times the threshold right before the downtime occurred. It is also quite encouraging that there were not any false alarms, in that the health value was below the threshold for all the remaining data samples, and once it crossed the threshold it never went below the threshold.

Figure 77: Regime Monitoring Map Based on Residual Clustering Results

In addition to calculating a health value, it is also of interest to understand which operating regime is best for monitoring the health condition of the wind turbine rotor shaft. The rotor degradation signature could be more pronounced under certain rotational speed or loading conditions, and the cluster results can provide an indication of which regime appears to be the best for observing this degradation signature. Using the indices from the selected cluster during each health value calculation, one would have the selected regime in terms of the rotational
speed and power output and this could be compared with the total number of samples. Effectively one could compare the number of selected instances in each power and speed bin, with the total number of instances in that respective power and speed bin. This proportion of selected instances can then be displayed as a 2-dimensional map (regime map), in which the bins that are selected more often would be the ones that are most conducive for detecting this rotor shaft problem. The regime monitoring map for this rotor shaft problem is provided in Figure 77, and one can observe that the best regime appears to be a generator speed between 840 and 960 rpm, and a power output between 300 and 600 KW. However, there are other bins that also seem quite suitable, and thus more historical cases would need to be accumulated before one could use the regime monitoring map to simplify the monitoring system and focus on a smaller set of operating conditions.

7.3.3 Benchmarking with Other Multi-regime Health Methods

Although the residual clustering method provided a clear health trend and an early detection of the rotor problem, it is still worthwhile to investigate how other multi-regime health monitoring methods would perform for the same data set. Modeling the feature distribution using a mixture of Gaussians and then comparing the overlap between the baseline distribution and the current distribution is a recent method that has been used for compressor condition monitoring [77] and also for wind turbine performance modeling [78]. The initial step in this procedure is to model the baseline feature distribution, \( H(X) \), by a Gaussian mixture model (GMM). A mixture model is shown in Equation 60, in which \( R \) is the number of mixtures, \( \rho_i \) is the weight of the \( i^{th} \) mixture, and \( \mu_i \) and \( \sigma_i \) are the mean and standard deviation for the \( i^{th} \) mixture. The feature distribution of the monitored system, \( G(X) \), is modeled in a similar manner [130].
\[ H(x) = \sum_{i=1}^{n} p_i N(\mu_i, \sigma_i) \]  

The comparison between the two distributions is computed using the L$_2$ distance, and the quantity is normalized by the L$_2$ distance value for each distribution as shown in Equation 61.

\[ CV = 1 - \frac{\left\| H(x) - G(x) \right\|_{L^2}}{\sqrt{\int \left( (H(x))^2 + (G(x))^2 \right) dx}} \]  

(61)

The L$_2$ distance between two distributions can be formulated using the integral expression in Equation 62, in which the expanded form of that expression is provided in Equation 63.

\[ \left\| H(x) - G(x) \right\|_{L^2} = \sqrt{\int \left( (H(x) - G(x))^2 \right) dx} \]  

(62)

\[ \left\| H(x) - G(x) \right\|_{L^2} = \sqrt{\int \left( (H(x))^2 + (G(x))^2 - 2H(x)G(x) \right) dx} \]  

(63)

Effectively, there are three integrals that need to be evaluated in Equation 63, but they all involve a multiplication between two Gaussian distributions. Thus, the first step is to compute the product between two Gaussians as shown in Equation 64, in which the product is also a Gaussian distribution multiplied by a constant $z_c$. The expression for $z_c$ is provided in Equation 65 and is based on the covariance ($\Sigma_a$ and $\Sigma_b$) and mean ($\mu_a$ and $\mu_b$) of the two distributions.

\[ H(x)G(x) = M(x) = z_c N(\mu_c, \Sigma_c) \]  

(64)

\[ z_c = \left[ \det(2\pi(\Sigma_a + \Sigma_b)) \right]^{1/2} \exp \left( (\mu_a - \mu_b)^T (\Sigma_a + \Sigma_b)^{-1} (\mu_a - \mu_b) \right) \]  

(65)

If one considers the original integral expression, the evaluation of this integral is actually equal to the constant $z_c$, since the integration of any probability density function (in this case a Gaussian distribution) for the entire support is equal to 1.
Thus, the overlap health value calculation can be carried out by estimating the parameters of the mixture distribution and calculating that constant value \( (z_c) \) using the mean and covariance of the two distributions. Using the previously described GMM health assessment method, a Gaussian mixture mode with two mixtures was used to model the baseline feature distribution. The current distribution was modeled as a single Gaussian, and this was then compared to the baseline feature distribution when performing the overlap calculation. The same feature set and baseline sample instances that were used for training the auto-associative neural network were also used by the GMM method. In addition, the threshold was set using Chebyshev's inequality, in which the same false alarm rate (0.0625) was used.

\[
\int_{-\infty}^{\infty} H(x)G(x)dx = z_c N(\mu_c, \Sigma_c) = z_c
\]  \hspace{1cm} (66)

Figure 78: Gaussian Mixture Model Overlap Health Value for Wind Turbine Rotor
The health assessment results using the GMM model are provided in Figure 78, in which one can observe a rather noticeable indication of the problem well in advanced of the maintenance event. In addition, none of the recordings after the maintenance event are above the threshold and thus there are not any instances of a false alarm. However, there is some variation in the health value over time and it has less of a clear increasing trend when compared with the results from the residual clustering method.

The use of principal component monitoring statistics, $T^2$ and the square prediction error (SPE) are two of the more established methods for assessing the health condition of a system or component. The interested reader can refer to a recent article by the author [52] for more details on the principal component health monitoring method. The direct application of the PCA monitoring method is only suitable for systems that operated in a single operating regime. However, the use of multiple local models, in which there is a baseline health model in each regime, can be used to overcome the multi-regime aspect. For this application, the generator rotational speed was used as the operating parameter, and two local baseline models were constructed. A model was based on baseline data in which the generator speed was less than 900 rpm and another model was based on baseline data when the generator speed was equal to or greater than 900 rpm. For calculating the health value, the $T^2$ distance metric is calculated using the expression provided in Equation 67, in which $\{u\}$ is the data projected in the principal component space and $\Sigma$ is the eigenvalue matrix. Effectively, one would use the appropriate health model and distance calculation based on whether the generator speed was below or above 900 rpm.

$$T^2 = \left\{u\right\}^T \Sigma^{-1} \left\{u\right\}^T$$  \hspace{1cm} (67)

The principal component monitoring method has established statistical thresholds for the $T^2$ health value, and the same alpha level (0.0625) was used so it could be compared with the
other health monitoring methods. In a similar manner, the health value was normalized by the threshold, so any value above one would represent a degraded condition. The results of the principal component $T^2$ monitoring method is provided in Figure 79. One can observe a rather clear increasing trend in the health value. However, there is less of an advanced notice of the problem, since it crosses the threshold 14 days before the downtime event, while the residual clustering method provided 26 days of advanced warning.

![Turbine 1 PCA Health Value](image)

**Figure 79: Principal Component Analysis $T^2$ Health Value for Wind Turbine Rotor**

For further benchmarking, a popular health assessment algorithm based on a self-organizing map was used. The self-organizing map minimum quantization error (MQE) method has been used in a variety of case studies and provides a rather robust distance based health calculation. A flow chart of the self-organizing map (MQE) health calculation is provided in Figure 80, in which the first step is to train the self-organizing map with baseline data. It
effectively clusters the baseline data, and then when new data comes in, it compares the test feature vector to all the training vectors and finds the closest location on the trained self-organizing map. This training data sample that matches the test sample is called the best matching unit (BMU), and the Euclidean distance between the BMU and the current situation is denoted as the minimum quantization error (MQE). As one can observe, the health value is a distance value to the nearest training sample as opposed to a distance to a cluster center. More specifics and case studies on the application of the self-organizing map for health monitoring can be found in [141-142].

Figure 80: Algorithm Flow Chart of Self-Organizing Map Minimum Quantization Error (SOM-MQE) Health Monitoring Method

The self-organizing map was trained using the same baseline data set that was used by the other three methods (AANN, GMM, PCA), and the threshold was also set using Chebyshev’s inequality and an alpha level of 0.0625. In addition, the health value was normalized by the
threshold, so any health value above 1 would be considered above the alarm level. The results of the self-organizing map (SOM-MQE) health assessment method are presented in Figure 81. The health monitoring results using the SOM method are very promising, with a consistent and increasing health trend until the downtime occurs. In addition, the method is providing an advanced warning of the potential rotor problem 24 days in advanced, which is quite similar to the residual clustering method that provided 26 days of advanced notice.

![Turbine 1 MQE Health Value](image)

**Figure 81: Self-Organizing Map-MQE Health Value for Wind Turbine Rotor**

7.4 Comparison of Methods

The previous section presented the health assessment results for four different methods, and qualitative observations about the health trend were discussed. However, providing a quantitative comparison between the health assessment methods is less straightforward. There
are an abundant of metrics for comparing fault detection algorithms using a confusion matrix, receiver operator curves, and false positive and false negative rates. However, there are less established metrics for comparing health monitoring trends in terms of how clear and consistent of a trend it is compared to others. One of the recently published health assessment trend metrics is the monotonic metric, which determines whether the health trend is consistently increasing or decreasing. For robust health estimation and also to provide a better input into a prediction method, it is desirable to have a health trend that is monotonic and does not vary with positive and negative direction changes. Coble [5] defined a monotonic metric using the expression provided in Equation 68. The monotonic metric is based on the number of times the health trend derivative is positive and also how many times the health trend is negative, and is further normalized by the number of data samples. Effectively one would obtain a monotonic value of zero if the health trend slope was positive the same number of times it was negative, however if the slope was positive for all the data samples one would obtain a monotonic metric value of one. Thus, the monotonic metric value is bounded between zero and one, and values closer to one are desired while values closer to zero are indicative of a less desirable health trend.

\[
\text{Monotonic Metric} = \text{mean} \left( \frac{\# pos \frac{d}{dx} - \# neg \frac{d}{dx}}{n-1} \right)
\]

(68)

For carrying out the monotonic metric calculation, it is suggested to using a moving average to first smooth the health trend. In addition, a local regression is calculated on the derivative of the smoothed health trend, so that the metric is based on noticeable changes in slope of the health trend and not noise. For this study, an exponential moving average was used with a smoothing constant of 0.1 and a local linear regression was performed using a moving window of 5 samples. The results of the monotonic metric for the four health assessment methods are
presented in Figure 82, in which one can observe that the residual clustering method has the best performance. The self-organizing map had the second most monotonic health trend, while the PCA and GMM methods had a much lower monotonic value for their respective health trends. This highlights that the residual clustering method provided a rather consistent and clear trend in the wind turbine rotor health state, and a prediction algorithm could have further used this health input for forecasting and determining when to schedule the maintenance activity.

![Comparing Health Assessment Methods](image)

Figure 82: Monotonic Metric for Health Trend - Comparing Results from Four Different Methods

Another aspect to consider when comparing the health assessment methods is how early of a warning each method provided before the downtime occurred. Providing advanced warning of a problem allows one to schedule the maintenance when the power generation is not needed and there would be less financial impacts. A summary of the advanced notice that each of the
four health assessment methods provided is shown in Table 21. Surprisingly, the GMM health monitoring method provided the most advanced warning of the problem (33 days), however it also had a health trend with a low monotonic value. However, the residual clustering method provided an advanced warning of 26 days which was second best, but also had the most monotonic health trend. The PCA method only detected the problem 14 days before the downtime occurred, which is still 2 weeks, but not as early as the other three methods. The residual clustering method performed as well and in some aspects better than the other three health assessment methods, highlighting that the method is quite appropriate for multi-regime health monitoring applications.

Table 21: Comparing Health Monitoring Methods on the Basis of Early Detection

<table>
<thead>
<tr>
<th>Method</th>
<th>Time When Health Value Exceeds Threshold</th>
<th>Advanced Warning (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AANN FOM</td>
<td>4/20/2011</td>
<td>26</td>
</tr>
<tr>
<td>SOM-MQE</td>
<td>4/22/2011</td>
<td>24</td>
</tr>
<tr>
<td>GMM</td>
<td>4/13/2011</td>
<td>33</td>
</tr>
<tr>
<td>PCA - $T^2$</td>
<td>5/2/2011</td>
<td>14</td>
</tr>
</tbody>
</table>

7.5 Discussion of Results

The results from this offshore wind turbine condition monitoring study provided some key conclusions. The results from the data pre-processing and feature extraction step, highlight the use of a vibration data validation check to remove erroneous recordings, and also the use of a kurtosis based filtering method that can provide features that are more suited for detecting the problem at an earlier stage. In addition, this case study further demonstrates that the proposed residual clustering health monitoring method is uniquely suited for assessing the health condition of a multi-regime system. The residual clustering method provided the most accurate result for the sensor health case study (among 24 participants), provided more discrimination for
the eccentric gear and shaft imbalance faults in the gearbox case study, and also provided the most monotonic health trend for the offshore wind turbine rotor problem. There is significant evidence that this particular approach is well suited for monitoring multi-regime systems and provides another algorithm for handling these monitoring applications. Although it is unlikely that the residual clustering approach would outperform the other methods for every monitoring scenario, the results indicate that the method is fundamentally sound and should be considered if the signals or features have a nominal amount of correlation.
8.1 Conclusions

The focus of this dissertation was to develop a method that could accurately monitor the health condition of a multi-regime system. From reviewing the prior work, it was noted that there are two main aspects that should be accounted for when monitoring a system that operates in multiple operating conditions; the signal features are highly influenced by the operating conditions and the degradation signature is usually more pronounced under certain operating regimes. The literature review highlighted that conventional multi-regime approaches that use local models or distribution methods are not particular suited for handling both of these aspects. Thus, a residual clustering framework was proposed to properly model and normalize the effect of the operating variables and also to perform cluster selection to find the most conducive regime for observing the system degradation.

The residual clustering approach effectively involves a residual processing step, a clustering and cluster selection computation, and also a figure of merit health value calculation. For residual processing, there are a variety of methods that can be used, including regression models, linear principal component analysis models, neural network models, and auto-associative neural network models. Based on the simulated correlation studies, it was noted that the auto-associative neural network model provided the most appropriate modeling technique with the best detection and false alarm rate. In addition, for a set of baseline wind speed data from a metrological tower, the auto-associative neural network had the best prediction accuracy when compared with the regression, neural network, and linear principal component analysis methods. Although the auto-associative neural network had the best
prediction error for the wind speed data, it also had the most variation from run to run since there is more variation in training the auto-associative neural network when compared with the regression and principal component analysis methods.

The residual clustering approach was demonstrated in three case studies, with the first dealing with a wind energy anemometer sensor health application. The data was from the 2011 Prognostics and Health Management Society Data Challenge, in which the residual clustering approach provided the best detection accuracy when compared with the other methods from the 23 participants in the contest. Considering the variation in several operating variables such as the ambient temperature or wind direction, the auto-associative model was quite appropriate for modeling this situation and normalizing the effect of the operating variables. The anemometer friction based failure mode was also more pronounced under certain operating conditions and the clustering method and figure of merit health calculation were well suited for handling this aspect.

The results from the gearbox condition monitoring case study further demonstrated the technical feasibility of the residual clustering method. The residual clustering approach was amended to include an additional feature extraction step that used domain specific signal processing and feature extraction methods for gear and shaft components. After extracting the gear and shaft features, the residual clustering approach provided accurate detection results of the eccentric gear and shaft unbalance problems. In addition, the residual clustering approach provided more discrimination between the normal and degraded cases when compared with the Gaussian mixture model overlap method. The gearbox case study also examined the effect of using different clustering algorithms with regards to the detection accuracy of the residual clustering approach. The use of different clustering algorithms did not seem to have much influence on the detection results, however k-means had slightly better performance for this
gearbox data set when compared with hierarchical clustering and a density based clustering method.

A third case study involving a wind turbine drivetrain condition monitoring application was used to further illustrate the merits of the residual clustering method. This particular application was for an offshore 3MW wind turbine, in which the drivetrain is subjected to varying rotational speeds and torque loads. Time statistics, kurtosis filtering methods, and envelope analysis were used to process the vibration data, and a signal validation method was used to remove erroneous data files. The signal features were fused by the residual clustering method in order to monitor the health condition of the wind turbine rotor shaft. The residual clustering health monitoring approach provided an early detection of the rotor problem and a very noticeable trend in the health value until the downtime event occurred. The residual clustering method also provided the most monotonic health trend when compared with the self-organizing map MQE method, the local principal component analysis method, and the mixture distribution overlap method. The three case studies highlight that the residual clustering approach is fundamentally sound and should be considered along with other algorithms for monitoring multi-regime systems.

8.2 Future Work

The residual clustering approach for monitoring multi-regime systems was focused on the health estimation aspect. For multi-regime systems such as wind turbine drivetrains, mining equipment, aircraft engines, manufacturing assets such as machine tools or industrial robots, determining the health state is only one facet. Providing root-cause information on the particular failure mode that is occurring would also be beneficial when monitoring these multi-regime systems. Also, one would like not only an assessment of the assets current health but also an estimate of the remaining hours the component or system can operate until the appropriate maintenance action should be taken. A list of potential research directions for the extension of
this work is provided; with the idea that future researchers would consider these suggested directions for their own dissertation or thesis work.

a) For monitoring the degradation of a component or system over time, an appropriate threshold is needed, so one knows whether the problem is severe enough to take action. Chebyshev’s inequality was used in this dissertation work for setting the threshold for the calculated health value from the residual clustering method. Although Chebyshev’s inequality is applicable for any health value distribution, more sophisticated approaches could be considered in future work. One option would be to use Monte Carlo simulation for estimating the health value distribution and determining the threshold based on a defined false alarm rate. This type of Monte Carlo approach has been considered for simpler health value calculations such as a summation of several features. However, its implementation would be more difficult since the residual clustering health calculation is more complex and it is dependent on several variables, such as the correlation in the features, the amount of baseline training samples, and the subset of regimes the degradation signature is present in.

b) The aspect of variable or feature selection was not considered during this dissertation work; however it would be beneficial in a variety of applications in which the number of signals or features can be quite large. In particular, for semiconductor manufacturing, the number of signals can exceed 100 and the number of calculated features could be well over 1000. The primary assumption of the residual clustering method is that only baseline data is used for training the model. Thus any type of variable selection method would have to be based on the baseline data. The variable selection could be based on a wrapper approach that uses auto-associative neural networks or other residual processing models and iteratively adds or removes features based on the prediction accuracy (root-mean square error). This would
select features on the premise of modeling the nominal correlation relationship in the
data, but would not select them on the basis of detecting the different failure modes
that could occur in the system or component. Also, an initial feature removal routine
might also be necessary to remove constant features or features that have little to no
correlation with the other features.

c) Although assessing the health condition is a key aspect for condition monitoring, in
many instances, root-cause and diagnosis information is needed by the maintenance
technicians and operators to reduce the maintenance and repair time. The
dissertation work clearly demonstrated the feasibility of the residual clustering
approach for assessing the health condition of a multi-regime system, however root-
cause or diagnosis was not considered in this study. Considering that only baseline
data is provided for training the residual processing algorithm, supervised
classification methods could not be used for diagnosing the different component
failure modes. A more suitable approach would be to use an adaptive diagnosis
approach that learns over time and accumulates different failure patterns over time
for diagnosing the system. Although this would not provide diagnosis functionality
immediately, after accumulating enough historical patterns over the course of several
years, the diagnosis method would be quite robust. The inputs to the algorithm could
be the residual features values and a K nearest neighbor method could be used for
the diagnosis.

d) Predicting the future health state and the remaining useful life of the monitored
component or system remains the most difficult task. This task is even more difficult
for multi-regime systems, since the health assessment is more difficult and also the
future loads and operating conditions can influence the components life. The work
performed during this dissertation addresses the multi-regime health assessment
portion and the results indicate that the residual clustering method can provide a clear monotonic health trend. There are already developed methods for prognostics that use the current and past health estimates; these include machine learning methods, regression techniques, and stochastic filtering methods. They however are primarily based on the assumption that the future load and operating conditions are similar to the current loading conditions. Thus the main challenge for multi-regime prognostics would be how to properly account for the uncertainty in the future loading conditions. The prognostic algorithm could use historical data and distribution methods to estimate the likelihood of the future operating conditions. Alternatively to predicting the future loading conditions, one could provide the user with different failure prediction estimates based on different operating scenarios. This prediction scenario approach might be more useful for decision making, since one could reduce the operating load or speed to extend the components life.
REFERENCES


17. G.A. Cherry, Methods for improving the reliability of semiconductor fault detection and diagnosis with principal component analysis, PhD Dissertation, The University of Texas at Austin, 2006.


41. E. Bechhoefer, D. He, and P. Dempsey, Gear health threshold setting based on a probability of false alarm, Annual Conference of the Prognostics and Health Management Society, 2011.


47. W. Yan and F. Xue, Jet engine gas path fault diagnosis using dynamic fusion of multiple classifiers, IEEE World Congress on Computational Intelligence (IJCNN), 2008.


87. A. Hess, L. Fila, Prognostics, from the need to reality – from the fleet users and PHM system designers/developers perspective, Proceedings of the IEEE Aerospace Conference, 2002.


98. G. Fung, A comprehensive overview of basic clustering algorithms, June 22, 2011.


102. V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault
detection and diagnosis Part I: Quantitative model-based methods, Computers and
Chemical Engineering 27 (2003), 293-311.

103. D. Capriglione, C. Liguroi, C. Pianese, A. Pietrosanto, On-line sensor fault detection,
isolation, and accommodation in automotive engines, IEEE Transactions on Instrumentation
and Measurement 52 (2003), 1182-1189.

104. R.J. Patton, Fault detection and diagnosis in aerospace systems using analytical

105. J.W. Hines, R.D. Garvey, Development and application of fault detectability performance
metrics for instrument calibration verification and anomaly detection, Journal of Pattern

106. E.L. Petersen, N.G. Mortensen, L. Landberg, J. Hujstrup, H.P. Frank, Wind power

107. S. Murakami, A. Mochida, S. Kato, Development of local area wind prediction system for
selecting suitable site for windmill, Journal of Wind Energy and Industrial Aerodynamics 22
(2003), 679-688.


109. J. Beltran, A. Llombart, J.J. Guerrero, Detection of nacelle anemometers faults in a wind
farm, Proceedings of the International Conference on Renewable Energies and Power
Quality, April 15-17, Valencia, Spain, 2009.

110. PHM Society 2011 Data Challenge Competition,
[http://www.phmsociety.org/competition/phm/11]

111. E. Hale, L. Fusina, M. Brower, Correction factors for NRG #40 anemometers potentially
affected by dry friction whip: Characterization, analysis, and validation, Wind Energy 15
(2012), 489-502.


121. PHM 2009 Data Challenge Competition, [http://www.phmsociety.org/competition/09].

122. PHM Society Public Data Sets, PHM 2009 Data Challenge Competition Labeled Dataset, [https://www.phmsociety.org/references/datasets].

129. Y. Yang, Tutorial on clustering, DKE Group, September 2, 2011.


