APPLICATION OF DATA MINING ALGORITHMS FOR THE IMPROVEMENT AND SYNTHESIS OF DIAGNOSTIC METRICS FOR ROTATING MACHINERY

by

Nicholas Goodman

Bachelor of Science
University of South Carolina, 2006

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mechanical Engineering College of Engineering and Computing University of South Carolina

2011

Accepted by:
Abdel Bayoumi, Advisor
Jeffery Morehouse, Committee Member
William Ranson, Committee Member
Jianjun Hu, Committee Member
Yong June Shin, Committee Member
Tim Mousseau, Dean of the Graduate School
DEDICATION

This work is dedicated to my soon-to-be wife, Wan, who has stood by my side and patiently supported me throughout these past few years. Your encouragement and assistance on all things, large and small, are greatly appreciated. Without you, none of this would be possible.

This work is also dedicated to the memory of Sergeant First Class Timothy Cook, whose years of dedication to our laboratory were pivotal in the establishment and sustainment of our research program. You were a great mentor, co-worker, and friend to us all, and we all owe you our deepest gratitude for your service to us and our country.
ACKNOWLEDGEMENTS

Projects such as these require collaborative efforts from many individuals and organizations, and the author would like to acknowledge several of those whose contributions are responsible for making this writing a success:

From the University of South Carolina: the Dissertation Advisory Committee, especially Abdel Bayoumi and Jianjun Hu; fellow CBM Graduate Research Assistants, especially Trevor Roebuck and Praveen Nooli; and the entire CBM operator team.

From the South Carolina Army National Guard: Major General Lester Eisner, CW5 (retired) Lemuell Grant, and the late Sergeant First Class Timothy Cook.

From the U.S. Army: AED, AMRDEC, Apache PMO, and the CBM Office; especially Jamie Allen, Damian Carr, Gail Cruce, Harold Hardee, Jon Keller, Jeremy Partain, Chris Smith, and Al Trausch.

From other individuals: Lance Antolick, Paula Dempsey, Lee Goodman, and all family and friends of the author who provided support and encouragement throughout the course of this work.
Abstract

Common applications of Condition-Based Maintenance utilize sensors mounted on mechanical components to diagnose failure conditions and incipient faults. The algorithms which process the raw data into diagnostic and prognostic indicators are typically derived from theoretical models, traditional signal processing metrics, or through trial-and-error observations. Condition monitoring devices are becoming increasingly common and are producing large volumes of data, yet relatively few studies have examined this real-world field dataset. Utilizing common data mining algorithms, an inferential study was performed to create diagnostic indicators which can distinguish healthy and faulted drive train components. Data was collected from multiple rotating components from several hundred rotorcraft and from a laboratory setup of the same components. Preconditioning filters for noise reduction and dimensionality reduction were performed, and it was found that for all components in the study, it was possible to represent complex vibration spectra with a relatively few number of attributes. Uniqueness among the individual articles within the sample population was also observed, and the implications of these findings are discussed. From the results of the preliminary investigation, classifier models were built on laboratory and field data to identify components which were normal, nearing failure, or failed. Multiple evaluations of the classifiers were performed, and a general approach to achieve data-mining derived condition monitoring is proposed.
Dissertation Overview

This dissertation begins by giving the reader an introduction to predictive maintenance and the basics of data mining in Chapters 1 and 2. Both topics are very broad fields, and it is recommended that the reader consult with the appropriate texts [1] and [2] to gain a more in-depth understanding of the subject if necessary. The following chapters, 3 and 4, give a basic demonstration of applying primitive data mining algorithms to condition monitoring data and discuss the implications of these preliminary findings. The next three chapters, 5 to 7, present a detailed study from a single laboratory experiment. Chapter 5 gives an in-depth description of the experimental setup and test history, while Chapters 6 and 7 study two different data mining approaches to the same experimental data. The final content chapter, 8, takes the lessons learned from the lab studies and applies it to data collected in the field. The major findings are summarized in Chapter 9.

Manuscript Chapters

Three of the chapters in this work (4, 5, and 6) were previously published and presented at various American Helicopter Society conferences and can be found in the appropriate conference proceedings publications [3],[4], and [5]. A footnote reference is supplied on the cover page of each of these manuscript chapters.
# CONTENTS

Dedication .................................................................................................................................................. ii
Acknowledgements ................................................................................................................................. iii
Abstract .................................................................................................................................................... iv
Preface ........................................................................................................................................................ v
Contents ................................................................................................................................................... vi
List of Tables ........................................................................................................................................... viii
List of Figures .......................................................................................................................................... x

Chapter 1  Introduction to Predictive Maintenance .............................................................................. 1
  Background .......................................................................................................................................... 1
  Introduction to Condition Monitoring ................................................................................................. 3
  Theoretical Framework of Condition Assessment ............................................................................... 9

Chapter 2  Introduction to Data Mining .............................................................................................. 11
  Data Mining Basics .............................................................................................................................. 11
  Challenges in Condition Monitoring ................................................................................................. 15
  Literature Review ............................................................................................................................... 17

Chapter 3  Preliminary Investigation .................................................................................................. 21
  Background .......................................................................................................................................... 21
  Field Data Sources .............................................................................................................................. 21
  Singular Value Decomposition of Power Spectra ............................................................................. 25
  Anomaly and Fault Detection ............................................................................................................. 31
  Compression Implications .................................................................................................................. 31

Chapter 4  Observations on Data Requirements ................................................................................... 32
  Introduction .......................................................................................................................................... 33
  Main Body ........................................................................................................................................... 34
  Conclusions ......................................................................................................................................... 47

Chapter 5  Experimental Setup and Observations .............................................................................. 49
  Introduction .......................................................................................................................................... 50
  Experiments and Findings .................................................................................................................... 54
Conclusions ............................................................................................... 64

Chapter 6 Analysis of Lab Experiment Data: Condition Indicators .............. 67
  Introduction ............................................................................................... 68
  Main Body ................................................................................................ 71
  Conclusions ............................................................................................... 82
  Acknowledgements ................................................................................... 84

Chapter 7 Analysis of Lab Experiment Data: Spectra ........................................ 85
  Overview ................................................................................................... 85
  Characteristics of Data Set ........................................................................ 85
  Classification of Source Article ............................................................... 91
  Classification of Fault Class .................................................................... 95

Chapter 8 Analysis of Field Data: Spectra ...................................................... 99
  Overview ................................................................................................... 99
  Baseline Dataset ........................................................................................ 99
  Faulted Dataset ...................................................................................... 107

Chapter 9 Conclusions and Recommendations ............................................. 112
  Discussion ............................................................................................... 112
  Conclusions ............................................................................................. 113
  Recommendations ................................................................................... 115
  Final Remarks .......................................................................................... 117

References .................................................................................................... 118
LIST OF TABLES

Table 3.1 - Mechanical characteristics of the studied gearbox types ......................... 22
Table 3.2 - Sensor and measurement scheme for each type of gearbox ....................... 22
Table 5.1 – Standard loading profile for a standard 4 hour test run ............................ 52
Table 6.1 – Random Forest classifier results for predicting aircraft tail number from CI data.......................................................................................................................................................... 73
Table 6.2 – Performance metrics of a KNN classifier which predicted gearbox life fraction from CI values .......................................................................................................................... 78
Table 7.1 – Source article prediction classifier performance results ......................... 94
Table 7.2 – Various classifier performance metrics when evaluated with cross-fold evaluation.......................................................................................................................................................... 95
Table 7.3 – Classifier performance results trained on Articles 2 and 3 and tested on Article 1 .......................................................................................................................................................... 97
Table 7.4 – Classifier performance results trained on Articles 1 and 3 and tested on Article 2 .......................................................................................................................................................... 97
Table 7.5 – Classifier performance results trained on Articles 1 and 2 and tested on Article 3 .......................................................................................................................................................... 98
Table 8.1 – Detailed specifications of gearboxes examined in this study .................... 100
Table 8.2 – Detailed specifications of the bearings examined in this study ............... 100
Table 8.3 – Receiver Operator Characteristic (ROC) area for the Euclidian distance metric ........................................................................................................................................................................................................... 106

Table 8.4 – Source gearbox classifier performance results ......................................................................................................................... 107

Table 8.5 – Source bearing classifier performance results ............................................................................................................................. 107

Table 8.6 – Information gain metrics for selected reduced dimension features for fault-tagged dataset ........................................................................................................................................................................................................... 109

Table 8.7 – Fault class classification results with 200 features vs. the optimized 4 features ........................................................................................................................................................................................................... 111

Table 8.8 – Fault class classification results with 200 features vs. the optimized 4 features ........................................................................................................................................................................................................... 111
LIST OF FIGURES

Figure 1.1 – Organization of maintenance methodologies ........................................... 2

Figure 1.3 – Theoretical process flow of a condition monitoring system ................ 10

Figure 3.1 – Full sample dataset of acceleration (top) and velocity (bottom) vibration spectra .......................................................... 24

Figure 3.2 – Vibration spectra with RMS values overlaid and sorted by RMS. Dashed lines denote samples which were removed in filtering........................................ 24

Figure 3.3 – Relative information content vibration spectra for the three gearbox types 27

Figure 3.4 – Residual error spectra from compression and decompression utilizing first 200 singular vectors ............................................................. 29

Figure 3.5 – Comparison of the residual error from SVD generalization for measurements supplied and excluded from the decomposition................................. 30

Figure 4.1 - Relative total file size of different condition monitoring datatypes......... 38

Figure 4.2 - Variance analysis of all condition indicators for the studied aircraft type.... 43

Figure 4.3 - A comparison of raw spectral data (a) to compressed and decompressed spectra (b).......................................................................................... 45

Figure 5.1- USC AH-64 Tail Rotor Drive Train test stand............................................. 51

Figure 5.2 - AH-64 tail rotor and intermediate gearbox assembly location .............. 53

Figure 5.3 - Temperature-time plots of a thermally unstable intermediate gearbox ....... 55

Figure 5.4- Changes in grease coloration when exposed to 300 °F temperatures........ 56
Figure 5.5 - Internal components of an AH-64 tail rotor gearbox ........................................ 57
Figure 5.6 – Seeding procedure utilized to induce an output seal leak............................. 58
Figure 5.7 - Article 1 (a) and Article 2 (b) input gears after catastrophic failure .......... 58
Figure 5.8- Grease mixing observed through the static mast window of Article 3 after (a) 120 minutes and (b) 145 minutes of operation ................................................................. 59
Figure 5.9– Article 1 gear mesh harmonic amplitudes in the final four days of testing... 60
Figure 5.10– Time frequency domain analysis of Article 1 lateral accelerometer as it approached failure............................................................................................................. 61
Figure 5.11- Temperatures of the three measured locations on Article 3 prior to fault seeding ......................................................................................................................... 62
Figure 5.12 – Temperature of Article 3 as it experienced rapid heating from gear mesh lubrication starvation .............................................................................................................. 62
Figure 5.13 - Infrared thermal imaging of Article 3 after main........................................ 63
Figure 5.14 - Ejected grease column observed on intermediate gearbox test article....... 64
Figure 6.1 – Comparison of traditional CI decision trees (top) with data mining fault classification methods (bottom) ............................................................................................................. 70
Figure 6.2 – Histogram illustrating the various amount of use of the onboard HUMS system by aircraft .................................................................................................................. 71
Figure 6.3 – Sensor locations monitored by the Vibration Monitoring Unit [26]......... 72
Figure 6.4 – Sample crossplot matrix showing dependent (blue) and independent (red) condition indicators ........................................................................................................... 73
Figure 6.5 – AH-64 tail rotor drive train test stand........................................................... 75
Figure 6.6 – Tail rotor gearbox teeth damage due to lubricant starvation experiments ... 75
Figure 6.7 – Timetable and record count for the test stand configurations which were examined. Articles 1-3 all failed due to gear mesh lubricant starvation.......................... 75
Figure 6.8 – Selected CIs from the crossplot matrix showing changing values as the test articles approached failure ............................................................................................................. 78
Figure 6.9 – Classifier error plot comparing predicted gearbox life fraction to actual life fraction .............................................................................................................................. 79
Figure 6.10 – A comparison of the classifier error plots when predicting Article 2 from Article 1 training data (left) and predicting Article 1 from Article 2 data (right) ........... 80
Figure 6.11 – Predicted gearbox life fraction of healthy test stand gearboxes (top), a gearbox with a different type of fault (middle), and aircraft gearboxes (bottom) ................. 81
Figure 7.4 – Complete filtered dataset for acceleration (top) and velocity (bottom) spectra for the vertical (left) and lateral (right) sensor locations ...................................................... 87
Figure 7.5 – RMS signal values for acceleration (top) and velocity (bottom) spectra for the vertical (left) and lateral (right) sensor locations ......................................................... 89
Figure 7.6 – Error fraction curves for the studied units of measure and sensor locations 90
Figure 7.7 – Approximated damage fraction used for classification training............... 91
Figure 7.8 – Normalized Euclidian distance matrix of the acceleration spectra for the lateral (upper left) and vertical (lower right) accelerometer locations ............................. 93
Figure 7.9– Normalized Euclidian distance matrix of the velocity spectra for the lateral (upper left) and vertical (lower right) accelerometer locations ........................................ 94
Figure 8.1 – Full spectrograms of gearbox (left) and bearing (right) datasets examined 102
Figure 8.2 – Euclidian distance matrices of the vibration amplitude spectra for the four gearbox (left) and bearing (right) datasets ........................................................................... 104
Figure 8.3 – Probability density distribution of the Euclidian distance between multiple measurements from the same article (red) and measurements from different articles (blue) ................................................................. 105

Figure 8.4 – Complete gearbox dataset with tagged failed and un-failed articles ........ 108

Figure 8.5 – Crossplot matrix showing separability of failed an un-failed gearboxes ... 110
CHAPTER 1
INTRODUCTION TO PREDICTIVE MAINTENANCE

Background

Nearly all mechanical systems are subject to some form of wear or breakdown, even when the system is being utilized within normal operating conditions. In situations where the design of a component cannot be changed to prevent all failures, it is necessary to perform maintenance actions to remain in or restore the device to an operational state. The term maintenance encompasses all activities including tests, measurements, inspections, replacements, and repairs which accomplish this goal, and a wide variety of philosophies exists on the correct application of maintenance[1].

Although there are differing views on the sub-categories of maintenance methodologies, this writing assumes the model suggested by Mobley [1], in which there are three main types of maintenance: continuous improvement, corrective, and preventative (Figure 1.1). Continuous improvement, or improvement maintenance, encompasses actions of modification and redesign which attempt to remove the most common causes of breakdowns. Corrective maintenance is the process of troubleshooting and repairing systems which have already failed or suffer from intermittent failures. Preventative maintenance tasks are intended to prevent a loss of operating condition, and can include inspections, adjustments, or replacements when required by the program.
The differing schemes of deploying preventative maintenance (PM) are further broken into three main categories: time-based, usage-based, and condition-based[1]. Time-driven maintenance actions are typically inspections or mandatory replacements which are often recommended by the equipment manufacturer, based upon historical observations, or set from qualifications testing. Generally, scheduled maintenance is applied conservatively, often resulting in inspections which rarely find a fault and the replacement of still functional components. Usage-based maintenance is somewhat more advanced and relies on operational parameters to dictate when repair or adjustments are needed. These can include performance indicators such as fuel efficiency, or operational limits such as over-speed or over-torque conditions. Condition-based maintenance or predictive maintenance utilizes sensors, algorithms, and thresholds to identify fault conditions within components and to predict the optimal time at which to perform a repair or replacement prior to failure. Ideally, one fundamental tenet of this methodology is the utilization of on-line continuous monitoring devices; however, in practice, condition assessment tools may only be used periodically or when a usage-based indicator warrants, somewhat blurring the lines between the preventative maintenance schemes.
The successful implementation of a preventative maintenance program almost always requires the application of all three methodologies and seeks to enhance the performance of a system by maximizing safety, minimizing operational costs, or increasing total output. Achieving this balance can be quite challenging and requires detailed knowledge of the design and functional characteristics of all components within the system. It is then possible to perform a Failure Modes, Effects, and Criticality Analysis (FMECA), which guides the recommendations for which maintenance schemes to use on individual components. It may be surprising to note that through the FMECA process there are situations in which scheduled mandatory replacements of a part may be more cost effective than the capital and overhead associated with a dedicated on-line condition monitoring device. As the complexity of a maintenance approach increases, so do the upfront costs required to sustain it, and therefore for a successful implementation it is desirable to reduce the deployment costs of condition monitoring systems while improving their overall accuracy and reliability.

**Introduction to Condition Monitoring**

Depending on the device being monitored, a wide variety of condition monitoring technologies and techniques are commercially available for use. The most common technologies include vibration monitoring, thermography, tribology, and ultrasonic measurements[1]. Vibration monitoring is most frequently utilized on rotating or reciprocating mechanical systems and requires fairly complex tools to analyze and understand the results correctly. Thermal imaging is often used to check drive system performance as well as electronic switch gear. Tribological measurements can be used to monitor debris particles within a lubricant or even determine the quality of the lubricant
itself. Ultrasonic sensors can be utilized in the detection of acoustic emissions from static structures or shock pulse phenomenon from rotating machinery.

*Vibration Monitoring Fundamentals*

All rotating and reciprocating machines produce vibration, which result for a variety of different reasons. Rotor imbalance produces centrifugal reaction forces which are synchronized with the running speed of the machine, and these periodic loads, when transferred to the non-rotating support assemblies, result in periodic deflections of the surrounding structure. Similar load impulse phenomena occur for reciprocating machines. Variations in the demand load or torque can also cause external reaction forces, particularly when there is angular misalignment between two coupled rotating shafts. The final type of vibration encountered is structural modal deflection of both the rotor and the stator assemblies. These free vibrations are excited from the energy dissipated from the previously mentioned forced vibrations. Since these types of vibrations are governed by mass and stiffness parameters, they tend to stay constant regardless of the rotational frequency of the system.

The net vibration resulting from all these contributing sources results in a complex profile of displacement, velocity, or acceleration in a 6 degree of freedom domain. Selecting the correct units of measure and sensor type typically is application-dependent, since each has its own range of frequencies which it is optimal at detecting. In practice, it is rare to see all possible directions of vibration being monitored, particularly uncommon is the direct measurement of angular vibrations. In most applications, sensing is either performed with a one or two axis transducer, and in the case of rotating machinery, these are likely oriented in a single radial, orthogonal radial, or radial and
axial direction. Since the true characteristics of the dynamics of the system being monitored are reduced to a significantly reduced domain, the resulting signals are often complex, and separating the sources of each signal component is often difficult.

Despite the inherent challenges of analyzing and interpreting such complex data, vibration analysis of rotating machines is a particularly widespread practice when compared to other condition monitoring technologies. In fact, it is common in industry for the term predictive maintenance to refer to this specific task exclusively. For each type of component that is monitored, such as a bearing or gearbox, a wide variety of algorithms and techniques have been developed to assess its condition. A brief description of several of these algorithms is given below, and a much more detailed explanation of these and other methods are given in [1] and [6].

**Discrete Time Series Processing and Metrics**

A single vibration acquisition consists of a series regularly sampled measurements known as a time trace or waveform, and the generalized name for this domain of non-continuous measurement is known as a discrete time series. For a given axis of measurement it is common to plot the collected vibration amplitudes against time, and for fairly simple signals, an analyst may be able to identify the various contributing vibration sources. For more complex cases raw time domain signals are too complex to interpret directly, require a large amount of storage, and seldom provide any insight into the condition of the part which is being monitored.

As a result of this, it is common to use a number of summary statistics to describe the general characteristics of a given waveform. The first of these is simply peak to peak
amplitude, most commonly applied to displacement measurements. For a given displacement waveform $s_t$, the peak-to-peak metric is defined simply as:

$$PPK = \max(s) - \min(s)$$

Since all periodic waveforms have a mean of zero, the general method of describing the average signal level of a measurement is through the root-mean square (RMS) amplitude. For a vibration waveform, $x_t$, with $n$ samples in any unit of measure the RMS amplitude is defined as:

$$\text{RMS} = \sqrt{\frac{\sum x_t^2}{n}}$$

It can be noted that for zero-mean data, that RMS is equal to the standard deviation of the amplitude distribution. Although it gives some indication of the overall vibration energy level, it does not necessarily convey the severity or peakedness of the vibration. For that, a metric known as kurtosis is often used, and although there are multiple standard definitions, a simplified expression for mean-centered data is given as:

$$kurt = \frac{\sum x_t^4}{(n - 1) \text{RMS}^4}$$

These summary statistics are quite effective for simple components with relatively small numbers of moving parts and steady-state operating conditions. In the absence of any complex condition algorithms, these statistical features can sometimes provide sufficient information for basic troubleshooting and fault indication. For complex machinery operating in a wide variety of conditions, however, more advanced approaches are needed.
Time Synchronous Averaging

Time synchronous averaging is a simplified form of a time series signal which displays the mean vibration amplitude with respect to an interval fraction between a periodic event, most often a complete revolution for rotating components. Formally, given a continuous signal, $x(t)$, with a specified averaging period $\tau$, on the closed domain of $t:(0,n\tau)$, where $n$ is an integer number of averaging periods, the time synchronous average is defined on the domain $t:(0,\tau)$ as:

$$TSA(t) = \frac{1}{n} \sum_{k=0}^{n-1} x(t + k\tau)$$

For discrete time domain data, there are a number of aliasing corrections which need to be taken into account if the averaging period is not an integer multiple of the sampling period. Some techniques to address this issue are point-wise interpolation, or variable subsample sizing.

Time synchronous averages are particularly useful gear diagnostics since it allows for non-synchronized vibration sources, such as bearings or dynamic loads to be averaged out, leaving behind a much reduced waveform characteristic of the average vibration in a given rotation. One application of this method is the identification of localized gear defects, which are indicated by irregular time synchronous waveform shapes.

Spectral Analysis

A much more informative and useful technique for understanding vibration signals is achieved by transforming the time series waveforms into the frequency domain. The most common method for this approach is through the Fourier and other Fourier-derived transforms. The definition for this transform on continuous domains is:
and the equivalent transform on the discrete domain, known as the discrete Fourier transform is:

\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi f t} dt \]

where \( N \) is the number of real data points in the discrete data set \( x \). It should be noted that both of these transforms can be applied to real and complex valued functions; however, typically in vibration analysis, the time domain signal is completely real valued. As a result, in the case of real-input DFT, the frequency domain result obeys the symmetry condition:

\[ X_f = X_{N-f}^* \]

where the star denotes the complex conjugate. This observation validates the requirement that reversible transformations preserve information content of the signal. A real valued waveform with \( N \) elements produces \( N/2 \) real and \( N/2 \) imaginary values in the frequency domain.

These complex-valued spectra contain both phase and amplitude information for each frequency component in the original time domain signal. Often, phase information is not used in the analysis of vibration, and either the magnitude is used alone, or the square of the magnitude, known as the energy spectral density is used:

\[ \Phi(f) = X(f) X^*(f) \]

One other important property of the DFT is that it obeys Parseval’s theorem which states:
Theoretical Framework of Condition Assessment

In order to relate vibration to the condition of a part, a collection of dimension reduction functions, known as condition indicator (CI) functions, are applied to the various domains to produce a set of scalar values. These functions are typically model-based and attempt to extract information about a particular physical process or event, such as shock impulses or ball spin frequency amplitude. The values of CI functions are then considered attributes, or features, to be supplied to a classifier, which examines all available information and assigns the monitored component a fault classification. The traditional method of classification is threshold-based nominalization, typically converting each numeric feature into a description such as “Good,” “Caution,” “Alarm,” or “Failed.” This process is known as diagnostics, and in the next Chapter more advanced methods of classification are described which attempt to enhance the accuracy and specificity of fault classifiers. Alternatively, other classifiers can be applied to the collected data or the derived diagnostic classes to predict corrective actions or remaining useful life in a process known as prognostics. A summary illustration of the entire condition monitoring framework from acquisition to prognostics is shown in Figure 1.2, below.

\[
\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{f=0}^{N-1} |X_f|^2
\]
In most applications of condition monitoring, the classification and identification of faults is still performed by a human analyst. Much ongoing research is being performed on the development and refinement of condition indicators and methods for building computerized diagnostic and prognostic classifiers are being explored.
CHAPTER 2
INTRODUCTION TO DATA MINING

Data Mining Basics

Definition of Data Mining

Data mining is formally defined as any process which automatically seeks to find useful information from large datasets \([2]\). Often the goal is to find previously unidentified relationships and trends between distinct input attributes without necessarily explaining the causality of the underlying behavior. From this discovered knowledge, data mining users may also make predictions about the outcome of future observations. Many of the methods employed in the field of data mining are closely related to inferential statistics and have common use the computer science field of machine learning.

Differentiating data mining from other similar analysis approaches is the general requirement that the source dataset meets any of the following criteria: 1) the number of records or instances is large; 2) the number of dimensions or attributes is high; 3) the data type is heterogeneous and complex; or 4) the data is fragmented and distributed over multiple locations. The paradigm for data mining also differs significantly from traditional statistical methods, since it does not test a hypothesis which is defined prior to analysis. Rather, it can be viewed as the automated testing of all arbitrary hypotheses,
often on data which was not collected during the course of a traditional experimental design.

This research utilizes three of the most common data mining tasks: dimensionality reduction, classification, and clustering.

**Dimensionality Reduction**

Dimension reduction is the process of expressing a given number of attributes into a small set, typically through coordinate transformation or feature deletion. A wide number of algorithms exist, each with their own specific application strengths and weaknesses, and often applied to numerical data types. If the reduction is reversible with no residual errors, the reduction is said to be lossless. More often, however, dimension reduction is irreversible, although in many cases the benefits of expressing the data in the simplified domain outweigh the loss in accuracy. Two of the most common reduction algorithms are the principal components analysis (PCA) and singular value decomposition (SVD), both of which are derived from linear algebra.

PCA is an application of linear algebra in which each attribute of the transformed set is expressed as the linear function of the original attributes. This is performed mathematically by the determination of the eigenvectors and eigenvalues of the covariance matrix of the original dataset. The resulting attributes are orthogonal to each other, with the resulting sum of the variance equal that of the original data set. In essence, it is a coordinate system rotation in which the new attributes are ranked in terms of their contribution to the data set variance.

This is mathematically expressed as:

\[ D' = DU \]
Where $D'$ is the transformed data, $D$ is the $m \times n$ original dataset such that each attribute column has a mean of zero, and $U$ is the matrix of eigenvectors such that for a given eigenvector, the corresponding eigenvalues are ordered such that:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$$

In cases where the variance is significantly represented in the first $k$ principal components, it is possible to define an $n \times k$ reduction matrix, $U^*$, which consists of only the first $k$ columns vectors of the original matrix. The resulting transformation projects $D$ onto a $k$ dimensional space which ideally contains most of the original variance.

Another related and more general approach to dimension reduction is known as support vector decomposition (SVD). Whereas PCA requires the original dataset to be mean-centered for each attribute, SVD will work on nearly any matrix. Essentially, SVD is a form of matrix factorization which is often written in the form:

$$D = U\Sigma V^T$$

where $D$ is an $m \times n$ matrix, $U$ is an $m \times m$ matrix, $\Sigma$ is an $m \times n$ diagonal matrix, and $V$ is an $n \times n$ matrix. Both $U$ and $V$ are orthonormal, and are termed the left and right singular vectors, respectively. The diagonal matrix is termed the singular value matrix and is sorted such that $\sigma_{i,i} > \sigma_{(i+1,i+1)}$. Often, the singular values decrease very rapidly, meaning a good approximation of $D$ is possible given only a small subset of the singular vectors. Applications of this property include data compression as well as noise filtering, and for the case of data mining, SVD allows for fairly complex feature space to be represented by significantly fewer attributes.
Classification

Classification is the name of any task which seeks to assign a summarized category or value to a set of observations. The field of data mining classifiers is fairly broad and includes many common algorithms suited for the categorization of nominal and numeric data sets. A thorough introduction into the topic can be found in [2]. As those authors note, classification refers specifically to labeling objects into a discrete class type, whereas methods to predict continuous numeric outputs are generally referred to as regression. The outcome or purpose of classifiers can either be descriptive modeling, in which previously acquired samples are categorized, or predictive modeling, in which previously acquired samples are used to assign a category to new observations.

Applied to predictive maintenance, it is the core strategy in converting a set of measurements into a fault diagnosis or recommended corrective action. A set of class-labeled data is sampled to produce a training set on which a learning algorithm produces a classification model. The model is then evaluated using a test set, which is usually separate from the training set, and various metrics such as classification accuracy, precision, and recall are examined. For descriptive modeling, the classifiers can then be used to identify features which are responsible for influencing the output class, whereas in predictive modeling, the model can be used without any direct understanding of its inner workings to classify new data samples.

This is an important aspect of data mining approaches in the field of condition monitoring. Whereas traditional approaches attempt to define metrics or algorithms to create features reflective of component condition, data mining takes a top-down approach, seeking to identify relationships or predictors, regardless of the underlying
cause or reason. Although it may be of academic interest to study the resulting classification models, in practice, the application of these models is significantly more relevant.

**Challenges in Condition Monitoring**

As mentioned previously, there are several challenges which need to be addressed in the deployment of any condition monitoring program: most notably, the failure mode probability and effect predictions and the development and refinement diagnostic indicators. The objectives of the proposed work will seek primarily to address systematic methodologies to address the latter; however, the results may have implications which may ultimately preclude the need for failure mode predictions in future systems.

The art of identifying every possible failure mode of a component and then predicting its dynamic properties such as progression rate to failure and the resulting impact on the system can be a difficult and even impossible task. For example, a gearbox may have more than twenty failure modes, the relative likelihood and criticality of which must be known or estimated. These factors are usage dependent, so when designing a new system it is unlikely that the equipment manufacturer would be able to provide this information to the end user. Further complicating the condition monitoring system design process is the fact that for each failure mode with a sufficient likelihood and criticality, a corresponding condition indicator and its probability of success must be predicted.

In general, there are three common tools which are utilized to improve the accuracy of the FMECA stage in PM planning: modeling, testing, and historical event analysis. Modeling of individual components or sub-components utilizes the known physics and dynamics of operation and applied knowledge of the fatigue properties of the
materials either through direct solution or simulation. This technique can also provide a baseline prediction of dynamic response to fault conditions, assisting in the development of condition indicators. Model-based failure dynamic predictions are typically only valid for components that are non-complex or have near steady-state operating conditions. In cases where the likely modes of failure are already known and it is desired to determine how those failures progress, component testing can be utilized to generate fault signatures. Often, complex failures are fatigue or probabilistically governed, resulting in long experiments which are often too costly to justify; however, there are certain systems such as rotorcraft in which it is still economically feasible to do so.

For large organizations with well-established maintenance programs, PM planning often relies on historical records to complete a FMECA. For large systems already in operation, robust maintenance management programs keep detailed logs of all failures which occur, allowing for failure probabilities and their effects to be determined directly. Furthermore, once condition monitoring sensors are deployed, it is possible to apply continuous improvement to the indicator algorithms by examining the data from failures that were not predicted and then creating diagnostics that could identify that failure in the future.

Although post-implementation refinement of condition monitoring systems is already in practice in many facilities, this process is far from optimized. The goal of this research is to apply data mining as the primary framework for system evaluation and knowledge discovery and to establish a generic process by which diagnostic and prognostic metrics can be defined and improved. The approach to be utilized will have
implications to any application of condition-based maintenance, and it will enable a more effective strategy of reaching an optimal balance of preventative maintenance schemes.

**Literature Review**

In recent years, several other attempts have been made to improve the diagnostic capability of condition monitoring systems including automated fault classification from training or simulated data. Research in this field typically examines one or both of the following aspects of diagnostics: the features used for classification and the methods and optimization of creating the classifiers. For rolling element bearings, much of the published literature examines both aspects; meanwhile for gearboxes, it is more likely to see investigators searching for improved signal processing techniques to generate fault-dependent features.

In 2003, work performed by Samanta and Al-Balushi demonstrated how artificial neural networks (ANN) were highly successful at identifying the difference between failed and normal bearings when only supplied with time-domain features\cite{7}. The input features examined in this study, which included measurements such as root mean square value, variance, and skewness, are often considered some of the most primitive attributes that can be obtained from a vibration signal, and these findings highlight the capabilities of ANN type classifiers. In 2006, work by Sugumaran et al began to examine an automated procedure for time-domain feature selection, in which a decision tree was utilized to select the optimal attributes to supply a proximal support vector machine for fault type classification\cite{8}. This study attempted to distinguish between four different bearing conditions with classifier efficiencies all reaching 75\% or greater.
More recent studies have moved away from time-domain features, and this is often done through some form of demodulation or wavelet analysis. Two studies from 2008 examined ANN-type classifiers applied to different demodulated signal features. The work performed by Yaguo Lei et al utilized an empirical mode decomposition method, and a combination of attributes from the time-domain and frequency-domain parameters was supplied to a neuro-fuzzy inference system to diagnose compound fault conditions\[9\]. Al-Raheem et al achieved similar results using a Laplace-wavelet transform and a genetic optimized ANN classifier\[10\].

In comparison to rolling element bearings, gearboxes are generally considered more difficult to diagnose simply due to their increasing complexity. Not surprisingly, most research related to gearboxes assumes a more advanced feature set, and several studies are devoted to investigating algorithms which can be supplied to classifiers. In 2008, Loutridis examined the principle of self-similarity in gearbox vibrations as a method of identifying localized defects on a gear tooth\[11\]. It was found that statistical attributes from self-similarity scaling functions trend well with tooth crack length, and it is likely such a technique could be supplied to a classifier in the future.

A much more common approach, utilized by three different authors is the discrete wavelet transform (DWT). A 2009 study performed by Wu and Hsu applied features extracted from DWT to a fuzzy-logic inference classifier and claimed to have fault identification rates greater than 96%\[12\]. Other studies have utilized DWT as a comparison for accelerometer measurements with ultrasonic sensors\[13\]. Similar to research in bearings, other diagnostic approaches have utilized ANN-type classifiers. The 2010 paper by Saravanan and Ramachandran applied an ANN classifier to DWT features
with a variety of fault and lubricant conditions[14]. Their study also examined the minimum neuron count to achieve a reliable classifier accuracy, which in this case exceeded 80%.

Another tool which may be beneficial to the current proposed work is genetic algorithms and programming. As early as 1992, Zhang and Roberts described how diagnostic rules could be evolved to identify faults in continuous stirred-tank reactor systems[15]. In literature, there are several common uses of genetic algorithms for fault diagnostics such as parameter optimization, feature selection, and rule generation. Samanta et al showed in 2003 the use of genetic algorithms to optimize the number of hidden nodes within an ANN classifier[16]. That same year, Chen et al described a strategy for fault classification problems which utilized genetic algorithms both as a preliminary feature selector and also as a means of generating neural networks[17]. The procedure was demonstrated again by Saxena and Saad in 2007[18]. A common issue with genetic algorithms is the computation time that can result in optimization problems if the initial population does not represent an adequate guess for the solution. Zhang and Randall suggested in 2009 that envelop analysis of rolling element bearings could be optimized with a genetic algorithm, with initial guesses being supplied by lower-resolution kurtosis values[19]. Similar techniques are being applied outside the field of rotating machinery, as shown from the 2009 writing by Fei and Zhang in which genetic algorithms were shown to be capable of optimizing support-vector machine parameters for fault diagnosis on an electric power transformer[20].

The review of the current scientific literature suggests a continued investigation into methods of streamlining fault diagnostics and ultimately component prognostics is
warranted. The proposed work will be an extension of these findings and will seek to generalize an approach for the selection, optimization, and application of fault classifiers.
CHAPTER 3
PRELIMINARY INVESTIGATION

Background

Most of the data which is analyzed in the research and development of condition monitoring systems is generated in controlled laboratory experiments, typically with a specific indicator and fault mode which is being tested. Recent changes in the costs of monitoring technology, data transmission, and storage media have resulted in a fairly large number of systems with permanently installed monitoring devices which collect data from machinery during its normal use. This type of field data offers a new opportunity for analysis and due to its large volume, irregularity, and high number of uncontrolled variables, it is a perfect candidate for the application of data mining algorithms. Several chapters in this work are dedicated to the topic of field data analysis, and this section examines the important practice of dimensionality reduction used to improve the performance of data mining classifiers and clustering algorithms later.

Field Data Sources

The vibration data investigated in this study come from three different types of single-stage spiral bevel gearboxes detailed in Table 3.1, below. For each type, measurements were collected at a fixed running speed and power level in accordance with the established PM program. Due to memory and storage limitations, the CM
device did not store the original time series signal, and only the power spectra are available for analysis.

Table 3.1 - Mechanical characteristics of the studied gearbox types

<table>
<thead>
<tr>
<th></th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of teeth (input/output)</td>
<td>31/66</td>
<td>37/49</td>
<td>22/57</td>
</tr>
<tr>
<td>Input shaft speed (RPM)</td>
<td>21160</td>
<td>4860</td>
<td>3670</td>
</tr>
<tr>
<td>Mesh frequency (Hz)</td>
<td>12910</td>
<td>2999</td>
<td>1346</td>
</tr>
<tr>
<td>Transmission Angle (degrees)</td>
<td>90</td>
<td>70</td>
<td>90</td>
</tr>
</tbody>
</table>

Each of the three gearboxes were monitored with a single-axis accelerometer with a unique combination of sensor type, mounting style, and sensing direction (Table 3.2). All gearbox types were sampled at 48 kHz for a duration of 0.341 seconds or 16384 samples during each acquisition, giving a resolution of approximately 3 Hz in the frequency domain. The dataset consists of a large number of specimens for each gearbox type, and an average of 10 to 12 acquisitions per specimen was included in the study (Table 3.2). The time interval between successive measurements for a given specimen varied in accordance with the PM program requirements. Other factors, such as operating temperature, component life, and specimen condition were not recorded.

Table 3.2 - Sensor and measurement scheme for each type of gearbox

<table>
<thead>
<tr>
<th></th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor Type</td>
<td>Dytran 3077A</td>
<td>Dytran 3062A</td>
<td>Dytran 3077A</td>
</tr>
<tr>
<td>Mounting Style</td>
<td>Bracket</td>
<td>Stud</td>
<td>Bracket</td>
</tr>
<tr>
<td>Mounting Direction (relative to input)</td>
<td>Axial</td>
<td>Radial</td>
<td>Radial</td>
</tr>
<tr>
<td>Sampling Rate (kHz)</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Number of Specimens</td>
<td>186</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>Number of Acquisitions</td>
<td>1886</td>
<td>1041</td>
<td>1041</td>
</tr>
<tr>
<td>Average acquisitions per specimen</td>
<td>10.8</td>
<td>11.7</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Since the sensors used to monitor the gearbox vibrations are accelerometers, the acquisition system represents the discrete vibration power spectra in units of acceleration.
Often in condition monitoring, it is also useful to express the vibration in units of velocity, so the original spectra were converted using the following formula:

\[ v_i = \frac{2n a_i}{f_s} \frac{1}{i} \]

Where \( v_i \) and \( a_i \) are the velocity and acceleration magnitudes at the discrete spectrum index, \( i \), respectively; \( n \) represents the number of frequency bins in the spectrum; and \( f_s \) is the sampling frequency. It can be observed from this equation that vibration spectra in units of velocity will have a higher concentration of energy in the lower frequencies when compared to acceleration spectra. Throughout this chapter, both cases will be considered.

An initial qualitative assessment of the datasets was performed by plotting a pseudo-spectrogram, in which the time axis has been replaced by sample index (Figure 3.1) and is ordered by specimen and then by acquisition time. Examination of these spectrograms reveals a general consistency between the various specimens and samples; however, all three article types contain samples of lower than normal and higher than normal amplitude. It is assumed that these extreme fluctuations may not the result of normal sample variance, but rather due to sensor error or a component fault. Since the objective of the analysis is to define a healthy gearbox model for each specimen type, a simple method for classifying and removing the outliers was needed.

Computing and sorting by the RMS of each sample shows that for all three gearbox types there is an approximately linear distribution of RMS magnitudes between the fifth and ninety fifth percentile (Figure 3.2) for both the acceleration and velocity spectra. A filtering rule was then established to reject all samples not within the specified inter-percentile range. It should be noted that due to the different relative distribution of
signal energy in the acceleration and velocity domain, the RMS percentile values were found to be independent between the two units of measurement, causing the rejected samples to be selected differently depending on which units were chosen.

Figure 3.1 – Full sample dataset of acceleration (top) and velocity (bottom) vibration spectra

Figure 3.2 – Vibration spectra with RMS values overlaid and sorted by RMS. Dashed lines denote samples which were removed in filtering
Singular Value Decomposition of Power Spectra

As explained in Chapter 2, singular value decomposition and the closely related principal components analysis are two common methods for dimensionality reduction. The vector of amplitudes which comprise each individual spectrum contains a large number of dependent attributes which may be greatly reduced by any of these common algorithms. For each gearbox type with a sample size \( n \), the number of discrete frequency bins, \( m \), per sample was a constant 8192, resulting in three \( n \times m \) matrices. It is assumed that there exists some number, \( l \ll m \), of abstract features such all significant components of the vibration spectra can be represented.

For datasets with a large number of attributes, performing a traditional PCA is computationally complex, requiring an eigenvalue solution for the \( m \times m \) covariance matrix, which may also be computationally expensive to determine. There exists several iterative partial least squares methods, such as the NIPALS algorithm [21], which address this problem, however, preliminary investigations into PCA techniques found that for these datasets, solution time was on the order of magnitude of hours of computation and required preprocessing techniques such a data centering and unit transformations to produce meaningful results. Also, the iterative methods such as NIPALS were not always successful at producing the correct order of principal components due to susceptibility of the algorithm on the initial guess conditions [22].

Singular value decomposition does not have these challenges; since it is a fairly common linear algebra technique, many software applications have built-in functions to perform this task. For data sets such as these gearbox vibration spectra, computation time is on the order of a few minutes, and it requires no pre-conditioning of the source data.
Due to these procedural advantages, the research holds that SVD is superior to PCA for the dimension reduction of vibration spectra. Additional benefits, such as interpretability of results, are explained in the following sections.

**Qualitative Assessment of Projection**

The Eckart-Young theorem states that dimensionality reduction by the SVD method is a solution to the general least squares approximation of a matrix projected to a lower rank. Furthermore, the $k^{th}$ diagonal element of the singular value matrix, $\Sigma$, is equal to the root-mean square of the approximation error. That is given some $m \times n$ data matrix $D$ which is projected by SVD to an $m \times k$ matrix and then mapped again to another matrix, $D^*_k$, in the original $m \times n$ space then

$$\|D - D^*_k\| = \sigma_{k+1}$$

Since the singular value matrix expresses the RMSE, it is possible to know how well a given SVD projection approximates the original dataset. By observing that $\|D\| = \sigma_1$, it is possible to define the relative error resulting from reducing a data matrix to rank $k$ as:

$$\frac{\text{RMSE}}{\text{RMS}} = \frac{\sigma_{k+1}}{\sigma_1}$$

One final consideration while performing dimension reduction on a data matrix is the effect of sample size on the maximum number of dimensions which can be obtained. Since the rank of an $m \times n$ matrix cannot exceed $\min(m, n)$, in situations where the number of instances is less than the number of attributes, the reduced attribute set will be limited by the number of samples in the data set. In order to fully understand the
goodness of fit a given projected space has to the underlying data structure, it is useful to compare the error distribution across multiple sample sizes.

An SVD was performed on the acceleration and velocity spectra of the three gearbox types, and by plotting the resulting relative error against all possible reduced dimension spaces, it is possible to observe that the error fraction diminishes rapidly within the first few hundred singular values (Figure 3.3). Recalling that the original power spectra consisted of 8192 attributes, this indicates that a significant reduction in dimensionality can occur with only a minimal residual error. Because the endpoint behavior on the full population curves indicated some sample-size effects, a second SVD was performed on a random 50% subsample of each gearbox type to better understand when these effects become noticeable.

![Figure 3.3 – Relative information content vibration spectra for the three gearbox types](image)

It can be seen from the figure, there is a strong overlap between the full sample and subsamples within the first 200 dimension for all three gearbox types. This implies that in this region, the error fraction curve has converged onto the full population characteristic curve, and that additional representative samples will have no effect on the SVD results. Also, it can be observed that in all cases, the velocity power spectra can be approximated with a fewer number of attributes than the acceleration spectra. The cause of this phenomenon is unknown, however, it is theorized that the conversion to the
velocity spectrum resulted in an overall loss of data complexity since it is known to attenuate higher frequencies which are more likely to contain high-entropy shock pulses and white noise. One final observation which can be made from these figures is the apparent decrease in signal complexity between Type I to II and Type II to III gearboxes, as evidenced by the reduced number of needed dimensions to attain a 1% error fraction, particularly on the velocity curves. Since all three gearboxes have a similar design, it is not believed to be related to the mechanical components within the gearbox, but rather may be indicative of the ratio of the rotational speed or gear mesh frequency to the sampling frequency.

Effects of Sample Selection

Since any dimension reduction scheme is dependent upon the given training data set, it is often of interest to examine the goodness of fit the projected model has on data on which it has not been trained. From the previous section it was shown that spectra from all three gearbox types had great potential for dimension reduction with a relatively small resulting error. Recalling that the data supplied to the SVD algorithm had already undergone filtering, one it is of interest to test the reduction scheme on the data which had been previously rejected. For this study, the reduced dimension count $k$ was set to 200, and the entire unfiltered dataset was compressed and decompressed using the following formula:

$$\text{The resulting spectral residuals can then be visualized with a spectrogram-style plot with the filtering boundaries shown below (Figure 3.4). For each acquisition the root mean squared error was also computed and plotted (Figure 3.4), and from this graph it can be observed that samples which had been rejected due to high RMS signal values}$$
also result in higher error when being projected to the lower dimensional model. Of particular note is the severity by which the error increases when moving from samples which supplied in the training set to those outside the training set. This may be indicative that the SVD method may be good at decomposing the given vibration spectra matrices; however, the results may not be describing a general model of each gearbox type, but rather represent a highly specialized space unique to that particular dataset.

Figure 3.4 – Residual error spectra from compression and decompression utilizing first 200 singular vectors
Since the intended goal of the dimensionality reduction is to create a set of parameters which describe a general case of a given vibration spectra, the effects of sample selection were explored further. The RMS-filtered datasets of each gearbox type were subdivided and reordered into an equally-sized training and test set by random sampling. An SVD was performed on the training set, and both sets were then compressed and decompressed by the using the first 200 singular vectors of the results. For each measurement, the error fraction was determined using the previously defined method and then plotted (Figure 3.5). It is immediately clear that for all gearbox types and units of measure, the attained error is noticeably higher for samples on which were not included in the training set.

This is indicative of the phenomenon known as over-fitting and means that for the selected number of reduced dimensions, in this case 200, some of the information in the reduced domain describes special characteristics of the supplied training set and not the generalized spectrum of a given component. One approach to solving this problem would
be to identify the ideal set of singular vectors which minimize the total residual error on the test set. In this case, however, the results indicate that the average residual error encountered in the test set is acceptably low, and further optimization of the vector selection is unnecessary.

**Anomaly and Fault Detection**

The findings of this initial assessment confirm dimensionality reduction by singular value decomposition is particularly effective on vibration power spectra from gearboxes run at constant speed and torque. It was shown that for three different gearbox types with different operating characteristics, all collected spectra could be described in a reduced space of less than 200 attributes with an average error of less than 5%. Later chapters will examine the efficacy of common classifier and clustering algorithms on these reduced spaces and compare their performance with features created from signal processing algorithms.

**Compression Implications**

A secondary implication of these findings is that vibration power spectra from these articles are highly compressible. In this study, an original space of 8192 features was reduced to 200, resulting in a compression ratio of greater than 40 to 1. Such an observation may have little relevance for laboratory testing, but could have significant impacts on fleets of condition monitoring systems which are deployed across a large number of decentralized locations. Chapter 4 explores one such case and discusses aspects of data storage and transmission which would be alleviated by the ability to compress and decompress vibration spectra in this manner.
CHAPTER 4
OBSERVATIONS ON DATA REQUIREMENTS

Introduction

Background

The effectiveness of condition-based maintenance (CBM) practices can be enhanced by centralizing relevant maintenance data and condition monitoring (CM) data and performing fleet-wide analyses to improve the performance of diagnostic and prognostic algorithms. This practice is often utilized on systems in which CBM technology has been deployed before appropriate thresholds of known condition indicators have been established or before all useful condition indicators have been identified. Data-driven enhancement is an appropriate alternative to laboratory or model-based enhancement when the operating environment of the monitored system is diverse or when it is expected that the non-optimized CBM system will still generate a return on investment.

For US Army Aviation, certainly both conditions are met, since its fleet of rotorcraft operates in extremely varied conditions and the relative costs of the onboard systems are far outweighed by the potential gains of utilizing component health monitoring. In order to maximize the utility of the recently adopted technology, the Army is currently investing in both laboratory-based component testing and a large data repository for fleet analysis known as the CBM Data Warehouse. The repository stores information on aircraft flights, maintenance histories, and data from onboard condition monitoring devices.

During the past several years, the size of the US Army CBM Data Warehouse has been growing at such a significant rate and cost as to bring into question its effectiveness at producing the desired results. Furthermore, the movement of such large volumes of
data often from deployed units overseas presents a pressing burden on Army telecommunications infrastructure. Little work has been done to quantify the functional efficiency of data centralization or to establish limits on the size of such a program.

Data under Study

Most of the numbers presented in this study derive from 445 acquisitions taken from a random sample of 89 Army rotorcraft. The measurements were all taken while the aircraft were on the ground with flat pitch at full running speed. Statistical summaries of the condition indicators were performed on all available data from this particular type of aircraft.

Main Body

CM Data Types and Prevalence

Prior to performing a full-scale evaluation of the data which is collected, it is useful to review commonly utilized data types in standard CBM practices [1].

Discrete Measurements

Discrete measurements are often performed on properties which are generally considered constant over small time intervals. Typical examples of such measurements would include speed, torque, temperature, strain, etc., and the sampling rate can range from a few times per second to one measurement every several minutes. On many of the currently utilized rotorcraft CM devices, the usage of discrete parameter measurements is somewhat limited. Due to the perceived superiority of vibration monitoring techniques, the speed of particular rotating components is often the only recorded discrete measurement.
**Usage Parameters**

A more recently adopted practice in the field of rotorcraft CBM is a technique known as usage monitoring, which classically is defined as a compliment to, but outside of the definition of condition monitoring. The various usage modes of a rotorcraft are typically defined in terms of flight regime, therefore, this technique may sometimes be called regime recognition. Once a particular flight regime is identified, the predicted loading parameters are assigned to each component within the aircraft. Fatigue theory, particularly Miner’s rule, is then applied to predict the life fraction consumption of a particular part. When a component reaches a predetermined maximum life, it is then removed from the aircraft. Due to its relatively recent adoption, usage monitoring presently represents a relatively small fraction of collected CM data.

**Time Series**

Time series data can refer to any measurement taken over uniformly spaced time interval; however, in the context of CM devices, it typically refers to vibration or acoustic signals. Vibration data in the form of raw accelerometer measurements represents a fundamental data type from which all other vibration-derived features are defined. Since it contains the most information, it tends to be difficult to interpret and has a large storage requirement. Therefore, time series data is usually collected, processed, and then discarded by a CM device in normal operations. In the case of Army rotorcraft, this type of data is stored only on rare occasion, and represents almost none of the current CM dataset.
**Frequency Spectra**

Frequency domain data is a derived data type created by performing a discrete fast-Fourier transform on time series data. Used on accelerometer measurements, it breaks a raw vibration signal into discrete frequency components which are easier to interpret. Technically the transformation is lossless, i.e., an $n$ sample time series array would produce two $n/2$ arrays, one representing component amplitude and the other, phase. In most vibration analysis, phase information is discarded, leaving behind a power spectrum which requires 50% less storage space than the original signal. Vibration spectra are considered one of the most useful tools for trained analysts and represent approximately 90% of the total data created for the studied rotorcraft platform in the US Army.

**Time Synchronous Averaging**

Time averaging is a technique used in vibration analysis to identify accelerations that occur in-phase with a periodic event. In this way, out-of-phase signal components are averaged out, producing a smoother and shorter time series signal. A common application for this technique is gear analysis, in which the phasing event is defined as one revolution of the input or output shaft of the gearbox. By averaging the signal from multiple sequential revolutions of the gear, individual teeth can be identified and local tooth defects found. The amount of data generated is equal to the product of the sampling frequency and the period selected timing event. In the current context, time synchronous averages constitute 7% of the total data generated.
**Condition Indicators**

The most concise form of CM data is a set of scalar features or attributes known as condition indicators (CIs) which are defined as a function of one of the other data types. Some examples are RMS, kurtosis, and skewness which are defined for time series data, and band-summation, band-averaging, and peak detection for frequency domain data. Sometimes condition indicators are defined as weighted sums of other condition indicators and normalized on a scale of 0 to 1. Often CIs are assigned threshold-based fault classes such as Normal, Caution, and Exceeded or with colors like green, yellow, and red, which make condition indicators relatively simple to interpret. The total amount of CI data generated depends on the number of identified CI functions, but it is significantly less than the source data types. Also, due to their scalar nature, individual files representing CIs may contain more header information such as data source and time information than actual CI values. Taking this into account, these useful metrics require a relatively small storage space and constitute less than 2% of all CM data for the studied fleet.

**CM Data Summary and Implications**

Summarizing the data collected for the studied sample aircraft, it becomes apparent that the primary contributing data types are vibration spectra and time-averaged vibration signals (Figure 4.1), which together represent almost 98% of the total generation and transmission. Depending on the intended function of this data, it is reasonable to question the usefulness and necessity of centralizing these particular data types. As previously mentioned, both vibration spectra and time averaged vibration signals either require an analyst to interpret or a condition indicator function in order to
identify a fault class. It is reasonable to assume that these data are underutilized at the top-level, and in most cases, critical maintenance decisions will be based on CIs alone.

If the intended function of these data is the scientific advancement of the condition monitoring technology, it may still be statistically justifiable to store such large amounts of low-level data. There is little evidence to suggest, however, that this information must be centralized in an expedited manner. In following sections, both of these issues will be addressed.

![Figure 4.1 - Relative total file size of different condition monitoring datatypes](image)

**CM Generation Rate and Storage Requirement**

In order to successfully plan and implement a data centralization and analysis program, it is necessary to have estimates for the bulk rate of data generation. Since it is known that time-series data contains the most information and has the largest storage requirement, an upper-limit estimate of the total data generation rate can be given as follows:

\[
\hat{D} = B_s f_s t_m n_c f_m N \hat{X}
\]

Where \(\hat{D}\) is the total data generation rate, \(B_s\) is the number of bytes per sample, \(f_s\) is the sampling frequency, \(t_m\) is duration of time each measurement is sampled, \(n_c\) is the
average number of channels used in each measurement, \( f_m \) is the frequency of performing measurements, \( N \) is the number of systems in the fleet, and \( \bar{X} \) is the usage rate of each system. Values of these parameters are explicitly defined for fully implemented CBM programs; however, this study will suggest reasonable upper-limit estimates for each.

**Bytes per Sample**

Typical analog to digital converters in laboratory environments as well as on onboard CM devices will produce 8-bit, 10-bit, 12-bit, or 16-bit measurements in the form of signed integer values. Often, the data acquisition software will then convert these raw integers into floating-point numbers which require at least 32 bits (4 bytes) of storage each [23]. It should be noted that the type casting operation does not increase the precision of the original measurement, even though the storage requirement has doubled. Another issue to address is the modern programming practice of using double-precision floating point numbers or other numeric data classes which increase the storage requirement without adding any precision. Furthermore, taking into consideration accelerometer precision limits, it is reasonable that the number of bytes per sample never exceed 4.

**Sampling Frequency**

Selecting the correct sampling frequency for a given application is a somewhat complex and often contentious issue. Traditional signal processing would dictate that the sampling frequency can be selected using the Nyquist-Shannon sampling theorem, in which the sampling frequency is defined as twice the highest component frequency in the signal. In recent years, however, the field of mechanical fault diagnostics has
demonstrated the usefulness of high-frequency signal components far beyond the highest rotating component frequencies. Other issues to consider are the linear response range and resonant frequencies of the accelerometers themselves.

As a result of these complexities, it is common practice to collect accelerometer data at relatively high sampling rates. Although typical rotating component frequencies are in the 0 to 5 kHz range, it is not uncommon to see vibration measurements sampled at 24 to 96 kHz. For this model, the value for $f_s$ is assumed to be 48 kHz.

**Number of Samples per Measurement**

Whereas the sampling rate determines the range of detectable frequencies, the duration of each measurement determines the frequency resolution in the spectrum. For example, in order to have a 1 Hz resolution in the frequency domain, a measurement length of 0.5 seconds is required. From the previously selected sampling frequency, that equates to 24,000 samples per measurement. Also, due to the inner-workings of the discrete fast Fourier transform, the number of samples is typically selected to be a power of 2. In this model, it is assumed that a frequency resolution of 1 Hz is sufficient, therefore the value of $t_m$ is defined as $2^{15}/48000 = 0.68267$ seconds.

**Number of Channels per Measurement**

The average number of channels involved in a given measurement is heavily application-specific. On many modern rotorcraft CM systems, the number of accelerometers is typically in the range of 10 to 30. Not all of these are required for every measurement task, for example, a rotor track and balance may only use 2 to 6 accelerometers. Since this model is attempting to define a worst-case scenario, however, it assumes that every measurement will acquire a constant 32 channels of data.
**Measurement Frequency**

How often measurements are needed is also challenging to justify. As previously discussed, most systems used a tiered measurement schema of infrequent high-fidelity measurements and frequent basic measurements. For the sake of simplicity, this model will assume a relatively high rate of measurement of 12 measurements per hour, or once every 5 minutes.

**Number of Systems**

The number of aircraft in a fleet depends heavily on the user. For the U.S. Army, the total size of its rotorcraft fleet is in the thousands, whereas private companies may have fewer than a dozen aircraft. Since this study is primarily interested in the characteristics of Army rotorcraft data generation, this model estimates the total number of aircraft to be 3000.

**Usage Rate of Each System**

Helicopters are known for their heavy maintenance requirement. Hours of maintenance per hour of flight time vary significantly from platform to platform, and in the case of US Army rotorcraft, exact aircraft readiness rates are considered sensitive. In practical situations, it is unlikely that a single rotorcraft would exceed 6 flight hours daily (approximately 2200 hours per year), and so this value is defined as the worst-case usage rate for this model.

**Total Data Generation Rate**

From the assumed worst case parameters the total generation rate can be evaluated as:

\[ D = 4 \cdot 48000 \cdot 0.6827 \cdot 32 \cdot 12 \cdot 3000 \cdot 6 \]
A data generation rate of approximately 843 GB per day, although seemingly large, is readily attainable on modern enterprise-class servers. It also should be noted that this rate represents an upper-limit or worst case scenario, since in practical applications raw time series data is not collected as frequently. A more realistic estimate of total data generation is likely to be less than 50% of the above calculated values when considering more appropriate sensor count, measurement rate, and operational readiness.

Statistical Aspects of the Data

Condition Indicators

Most condition indicators are defined such that only positive values are possible; therefore, it is unlikely that a given CI will follow a Normal distribution \([24]\). This also means simple Z-score and t-score methods of approximating the true population mean do not apply. Without performing a distribution analysis for every single CI in use, simpler techniques such as percentile ranking can be performed in order achieve the task of anomaly detection. Even so, a basic variance analysis should be carried out on each CI in order to rank the quality and meaningfulness of each feature.

One metric to examine is the coefficient of variation, which is defined as the ratio of the standard deviation to the mean. Examining the summary statistics of the entire aircraft population, a total of 170 different CIs were studied each with sample sizes ranging from 6000 – 375,000 acquisitions. The coefficient of variation was determined for each CI, and the results were sorted and plotted in Figure 4.2.
Figure 4.2 - Variance analysis of all condition indicators for the studied aircraft type

It was found that 135 of 170 CIs (79.4%) had CVs less than 1, which means their signal-to-noise, the reciprocal of CV, is greater than 1. For condition indicators whose CV > 1.5 (12%), the high degree of variance implies that the defining function is either highly susceptible to noise or that the dataset has several extreme outliers which are skewing the results. An appropriate follow-up step would be to remove the top and bottom percentile of each set and repeat the CV analysis. If the high coefficient of variation is not due to outliers, it may be appropriate to discontinue using that particular CI. Conversely, CIs with a very low CV indicate functions which may not be sufficiently sensitive to detect condition changes or may be tuned to detect a fault which has a very low rate of occurrence.

**Vibration Spectra**

For a healthy component which runs in relatively steady state, such as a rotorcraft drive train part during a known flight maneuver, the vibration power spectrum measured will have a high degree of repeatability. Variation between different acquisitions would likely be the result of signal noise or changes in environmental conditions, while variation between multiple aircraft could be attributed to component age or
manufacturing variance. In either case, with a large enough representative sample, it is possible to define a normal spectrum for a given component and operating condition.

Since it can be shown that many of the component amplitudes within the spectrum are strongly covariate, it becomes possible to define this normal spectrum as a weighted sum of multiple principle component spectra [2]. If the principle spectra are known a priori, a given measurement can be converted and represented strictly by the set of weighing factors. If the measurement is converted into fewer principle spectra than data points in the original set, then there will remain a residual error spectrum which can be used to describe how well a given measurement conforms to the normal spectrum definition. A by-product of this technique is that if the root mean square of the residual is less than an established error threshold, then it can be discarded, and depending on the number of principle spectra defined, this can lead to very large data compression ratios.

A demonstration of this technique is shown in Figure 4.3. A sample set of spectra from a single component across multiple aircraft is deconstructed into 12 principle components, the residual is discarded, and then spectra set reconstructed from the compressed data. The results of the compression and decompression are very similar to the original dataset, yet it only requires 0.3% of the storage space.

**Storage and Transmission Efficiency**

**Relational Databases**

Relational databases, in which all data in stored in a tabular structure, are the most common method for storing and analyzing very large amounts of data. Due to their internal structure, database files are relatively easily searched and records can be added or
deleted quickly [25]. The overhead associated with the encapsulating data types, however, often results in very large storage requirements compared to a flat file.

![Figure 4.3 - A comparison of raw spectral data (a) to compressed and decompressed spectra (b)](image)

Most databases require fixed width memory allocation for a given field, meaning that in the case of textual data or variable-length binary object data, the same amount of memory is required per row, regardless of the size of the data stored there. Furthermore, every tuple, or row, in a database is meant to represent an object within an unordered set. For most database systems available on the market, this means that having an ordered set of numbers requires either a separate index field or utilizing binary object fields. In the
former case, this requires extra memory for the newly defined field, and in the latter, custom processing applications must be created to interpret the data.

For most types of condition monitoring data, direct conversion into a relational database is a poor choice for efficient storage and analysis, particularly for time series, frequency domain, and time synchronous averaged data types. Storage overhead can reach as much as 90% for these data types, meaning that a 4 KB measurement could require up to 40 KB of file space.

In order to utilize the benefits of relational databases but avoid the overhead for this type of data, the most efficient method for storing CM data is through database indexed external files. In this approach, all relevant header information, such as acquisition date and time, is stored in the relational database, meanwhile, the numeric arrays which hold vibration signals and spectra are maintained in separate files. This technique still requires custom software to be written for analyzing the arrayed data types, but allows for the software development to exist outside of the database application.

**Limited Storage and Priority-Based Transmission**

Previously, it was discussed that more than 98% of the CM data stored from a particular fleet and measurement scheme of Army rotorcraft represents high-level, non-queryable data types. It was also mentioned that the required sample sizes in order to identify population-wide norms were vastly smaller than the amount of data retained. For archival purposes, retaining a copy of all data ever collected may be useful; however, for the purpose of improving the performance of these CM devices, it is reasonable to propose that only a fraction of the data needs to be centralized and queryable.
If it is desired that each aircraft have a representative sample of spectra and condition indicators, it is statistically justifiable to limit the number of samples to 60 per sensor per aircraft. From the previous case study of 3000 aircraft with 32 sensors, this equates to a limit of 703 GB of storage to fully represent the monitored components within the fleet. This is less than one day of data when compared to the previously defined worst-case data generation scenario, and the remaining data could either be archived or discarded if it was determined unnecessary.

Furthermore, it would be possible to reduce the amount of data transmitted if only anomalous spectra were transmitted for centralization. For example, for deployed units it may be more efficient to only transmit spectra whose corresponding CIs do not fall within the inter-quartile range. This would effectively reduce transmission traffic by 45%, and if desired the non-transmitted data could be uploaded to the database after the unit returns from deployment.

Conclusions

Most data collected from the studied set of Army rotorcraft is likely seldom utilized

Nearly 98% of the data which is collected from the studied set of CM devices is of a type which is unlikely to drive commander-level or top-level decisions or component watch lists. Typically maintenance decisions are driven by condition indicators which account for only 1.51% of the data moved and stored.

There is a reasonable limit to the amount of data generated by a fleet of rotorcraft

Taking a worst-case scenario in which it was desired to migrate raw vibration data from a fleet of 3000 aircraft with 32 sensors and 72 measurements per day, it was concluded that the fleet-wide data generation rate would not exceed 844 GB per day.
Over a 10 year period, this amounts to 2.9 PB, a reasonable size for an enterprise-class server.

*Poor relational schemas can result in unreasonable storage requirements*

With inappropriate storage of ordered set data, such as time-series or spectral data, database overhead could reach as much as 90%. This could push the decade total for data generation into the petabyte range. Externally linking to binary measurement files would give the benefits of relational schema while cutting down on overhead.

*Priority-based transmission could significantly reduce network bandwidth*

Assuming a significant quota of representative data is kept for each aircraft, there is no need to continually transmit measurement samples representative of a healthy component. By simply restricting the inter-quartile range of one data type, vibration power spectra, it is possible to reduce network bandwidth by up to 45%.
CHAPTER 5
EXPERIMENTAL SETUP AND OBSERVATIONS

Introduction

*CBM Component Testing at the University of South Carolina*

Over the past decade, the University of South Carolina (USC) has held a strong working relationship with the South Carolina Army National Guard (SCARNG). During the early days of the Vibration Management Enhancement Program (VMEP), USC played a key role in the development of early cost-benefits models which demonstrated the usefulness and effectiveness of onboard health monitoring systems for the SCARNG fleet. These efforts expanded into a fully matured CBM Research Center within the USC Department of Mechanical Engineering, which hosts several aircraft component test stands in support of current US Army CBM objectives.

Within the USC test facility is a complete AH-64 tail rotor drive train test stand (Figure 5.1), which is designed to facilitate a scientific understanding of aircraft component conditions as they relate to TAMMS-A inspections, vibration signals, health monitoring systems output, and other data. These observations are necessary for the development of comprehensive and accurate diagnosis algorithms and prognosis models. The testing apparatus is also capable of being modified to test new and existing drive train components of military and civilian aircraft, including the ARH-70, CH 47, and UH-60 drive trains.
The test stand emulates the complete tail rotor drive train from the main transmission tail rotor takeoff to the tail rotor swashplate assembly. All drive train parts on the test stand are actual aircraft hardware, and it is capable of handing shafts installed at the maximum allowable misalignment of over two degrees. The structure, instrumentation, data acquisition systems, and supporting hardware are in accordance with military standards, and the test stand’s two 800 horsepower motors are capable of exceeding 150% of the actual aircraft drive train loading. The test stand was designed and built to accommodate the use of multiple Health and Usage Monitoring Systems and is currently equipped with a Honeywell Modernized Signal Processing Unit (MSPU). Alongside USC’s own data acquisition system, the implementation of currently fielded aircraft equipment helps validate test stand results with data from actual airframes.

Testing Procedures and General Observations

As on the actual helicopter, the tail rotor drive train test stand is a constant-speed and dynamic-loading power transmission system. The tail rotor drive shafts are spun at 101% aircraft normal speed, which is 4863 RPM, throughout the duration of a single test
run, while the output motor changes its braking torques to produce specified load conditions, which match flight regimes as requested by the Army Engineering Directorate.

Due to memory storage limitations on the MSPU, each test run is approximately 4 hours in length, at which point the data is downloaded onto a Ground Station computer. All studies presented in this research utilized the same testing sequence as shown in Table 5.1, below.

<table>
<thead>
<tr>
<th>Output Power (hp approx.)</th>
<th>Duration (min approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>198</td>
<td>50</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>252</td>
<td>50</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>330</td>
<td>50</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

The MSPU collects data periodically throughout the run, and once per hour a more detailed data Survey is completed. The current standard practice is to complete a Survey in a condition known as FPG101, or flat-pitch ground at 101% speed. These surveys correspond to the 10 minute, 30 horsepower loading intervals. A steady-state transmission of power through a gearbox would cause a constant heat generation condition, which would reach equilibrium at some temperature. As a result of the required cyclic loading conditions, the gearboxes on the test stand undergo thermal transients immediately after a load change. Because of the heat capacity of the gearbox
assembly, these transients occur over a relatively long time domain, and result in distinguishable intervals on a temperature-time plot.

**AH-64 Tail Rotor Drive Train Gearboxes**

The two gearbox assemblies that make up the AH-64 tail rotor drive train are known as the intermediate and the tail rotor gearboxes (Figure 5.2). The gear ratios of each gearbox reduces the drivetrain speed in order to provide high torques to the tail rotor blades. Unlike many gearboxes at these power loads, both intermediate and tail rotor gearboxes utilize grease, rather than oil, as their sole lubricant.

![Figure 5.2 - AH-64 tail rotor and intermediate gearbox assembly location](image)

The particular grease for these components is designated NS-4405-FG and consists of an ester base with a Lithium soap thickener. Ester greases are generally known for their high pressure stability and high heat resistance, and are outperformed only by the more expensive perfluoropolyether greases. In this case, NS-4405-FG has an operating temperature range of -65 °F to 275 °F.
Historically, it has been observed that some of the most common maintenance faults for AH-64 gearboxes are related to leaking or ejected grease. Some of these issues present only an inconvenience to maintenance crews, while others require extensive maintenance procedures or part removals.

Through multiple studies of AH-64 tail rotor drive train components, USC has made numerous observations and discoveries relating to the behavior of the gearbox-grease systems which have significant impact on the maintenance of the aircraft. A review of these findings is presented in this paper.

Experiments and Findings

Intermediate Gearbox with Severe Thermal Transients

An intermediate gearbox which had been removed from an aircraft due to unidentified vibration signatures was installed on the USC test stand for further analysis. Initially, the gearbox indicated no abnormal behavior, and the vibrations which were generated were inconclusive. At the end of the sixth test run during the 330 hp load interval, however, an unexpected rapid change in gearbox temperature was observed (Figure 5.3). Although, the over-temp condition of 300 °F occurred, the stand was not stopped since the test run was near completion.
During the following run, near the beginning of the 252 hp load interval, a similar thermal excursion occurred, resulting in an emergency stop of the test stand due to the over-temp condition. The stand was then allowed to cool to ambient temperature and the run was resumed in the 252 hp load interval. Steady state temperature conditions had not yet occurred when the stand load profile was cycled to an FPG101 configuration. Almost immediately after the 330 hp load was applied, a third thermal event was observed, which also warranted the shutdown of the test and the removal of the gearbox as a test article. This gearbox would be the first of many in which rapid changes of temperature at increasingly cooler starting conditions was observed. The sudden change in slope on the temperature plot indicates that an additional heat source beyond the normal effects of gear meshing may have been present. Following its removal, the gearbox was disassembled and studied to identify the source of the problem. The teardown analysis findings were inconclusive and found no major damage to gear teeth surfaces or rolling element bearings. It is therefore theorized that the cause of the sudden heat generation...
was not due to mechanical phenomena such as friction or wear, but rather due to an exothermic chemical decomposition of the grease.

Further ongoing follow-up studies are being performed at USC to characterize NS-4405-FG when heated beyond its specified operating limits. In one experiment, unused grease samples were placed in a 300 °F oven at ambient air pressure for a period of four weeks. During that time, it was observed that the grease began to change color from tan to red and finally to black (Figure 5.4). In its final condition, it was further observed that the grease no longer maintained a uniform semi-solid viscosity, and had separated into solid and liquid regions.

![Figure 5.4- Changes in grease coloration when exposed to 300 °F temperatures](image-url)
Tail Rotor Gearbox with Leaking Output Seal

In a separate study, three AH-64 Tail Rotor Gearboxes were tested to identify the survivability of the output shaft assembly ball bearings in a seeded-fault leaking output seal condition. Motivating this study was the Army required practice of replacing Tail Rotor Gearboxes at the first sign of an output seal leak. This was due to the perception that the static mast of the gearbox was sealed from the main compartment, meaning that static mast grease levels could not be observed or serviced (Figure 5.5).

The output seals were seeded to represent a worst-case scenario leak for the gearboxes, so a large amount of seal material and the compression garter spring were removed (Figure 5.6). When the first attempted article was tested, it was unexpectedly found that following a sufficient ejection of grease from the output seal, the main compartment service levels were also low. Immediately following a servicing of the main compartment, the grease flow from the output seal increased in volume.
Since it was thought that this behavior was impossible, the gearbox was disqualified and three more gearboxes, identified as Articles 1, 2, and 3, were selected. Article 1 showed similar behavior, causing the test plan to be modified such that the main compartment would only be serviced with the correct amount of grease at the beginning of each article test life. For all three articles, it was observed that a persistent grease leak through the output seal resulted in a loss of lubricant in the main gear compartment. Consequently, this condition ultimately resulted in lubricant starvation on the gear meshing region and catastrophic gear tooth failures (Figure 5.7). For Articles 1 and 3, the failure was accompanied by large thermal excursions, while Article 2 incurred a broken input gear tooth.

Figure 5.6 – Seeding procedure utilized to induce an output seal leak

Figure 5.7 - Article 1 (a) and Article 2 (b) input gears after catastrophic failure
The experiment for Article 3 included a specially modified static mast, which allowed visual inspections to be made of the static mast grease levels. A special red dye was added to the main compartment, and prior to cutting the output seal, the test stand was run to observe any grease mixing. Within 120 minutes of starting the test stand, the static mast window began to show a slight red coloration and by 145 minutes, grease from the two compartments had become thoroughly mixed (Figure 5.8).

![Figure 5.8- Grease mixing observed through the static mast window of Article 3 after (a) 120 minutes and (b) 145 minutes of operation](image)

From this study, various vibration signatures were collected that give some insight into the effects of lubricant condition on component vibrations. Examining the first and second harmonics of the gear mesh frequencies of Article 1 as it approached failure (Figure 5.9) shows noticeably cyclic behavior, despite the fact that all data points were collected in the same loading condition, namely FPG101. It is theorized that the cyclic behavior is strongly related to the gearbox temperature at the time of the Survey, and that accompanying viscosity changes in the lubricant cause the vibration signals to be affected.
This is evidenced even more when utilizing joint time-frequency domain analysis of the vibration signals for the same gearbox (Figure 5.10). In the days leading to the failure, the cold-start conditions show significantly less high-frequency noise than two hours into the run. Such noise is often associated with shock energy and indicates that in a lubricant-deficient gearbox, high temperatures cause a final protective layer of grease to become thinner and less effective.

During the experiment, several key observations about the thermal characteristics of a low-lubricant tail rotor gearbox were also observed. Contrary to the predicted behavior that a loss of lubricant would result in elevated temperatures due to increased friction, it was found that the heat generation regions of the gearbox maintained a consistent temperature profile until the point of failure. On Article 3, temperatures were monitored at three different locations: the input duplex bearing, the output roller bearing, and through a modified service plug which allowed a thermocouple to be placed very close to the gear mesh location.
Comparing the temperature profiles of Article 3 in original fully serviced condition with its imminent-failure condition, it can be observed that two of the locations saw no significant changes in thermal behavior while the output roller bearing actually experienced a decrease in temperature (Figure 5.11 and Figure 5.12). It is theorized that this effect is due to the loss of a convective transfer mechanism which distributes gear mesh heat throughout the gearbox. It is predicted that gearboxes experiencing low-lubricant conditions will typically exhibit large thermal gradients, as illustrated by the use of infrared imagery (Figure 5.13).

As observed previously, Article 3 experienced three over-temp conditions each at successively cooler initial temperatures. Examining the first of these events, it is clear that the magnitude of the thermal gradients increases severely during the thermal transients, and that the gear mesh region can reach temperatures approximately 100 °F hotter than the highest monitored location on the actual aircraft (Figure 5.12). This indicates the possibility that in certain scenarios, a gearbox may reach temperatures
beyond the operational limits of the grease without indicating this information to the flight crew.

Figure 5.11 - Temperatures of the three measured locations on Article 3 prior to fault seeding

Figure 5.12 – Temperature of Article 3 as it experienced rapid heating from gear mesh lubrication starvation
One presently ongoing investigation by USC has been to characterize and identify the cause of intermediate gearbox grease ejections, which frequently occur through the breather port. In this study, two articles that were removed from actual aircraft for this behavior are being utilized.

The first step in characterizing this fault was to attach a transparent pipe to the breather port so that the static pressure head of the grease could be determined. During the first 4 hour test run, grease was ejected out of the breather to a maximum height of approximately 14 inches (Figure 5.14). Based upon the density of the grease at room temperature, this would indicate an internal gearbox pressure of approximately 0.5 psi. The source of such a relatively low pressure has not yet been verified, but may be the result of centrifugal fluid dynamics of the grease or by thermal expansion of trapped air within the gearbox.
Another interesting observation is that the grease ejection only occurred during the first day of testing. At the end of the day, the grease settled back into the gearbox, and since then no further ejections have been observed coming from the breather. It is theorized that this may be due to rapid changes in grease properties such as a decrease in viscosity, thus changing the dynamics of air pocket and foam generation. Future work will seek to quantify if such changes are, in fact, occurring and to determine what implications it may have on the operational practices of the intermediate gearbox. Other possible follow-up investigations may attempt to modify the breather such that grease ejections are less severe.

**Conclusions**

1. Repair of Leaking Output Seals

The most practical conclusion that results from the findings presented is that unit-level replacement or repair of AH-64 tail rotor gearbox output seals is feasible. Since grease has been observed to mix and transfer freely between the two gearbox compartments, there is no longer an immediate need to ground aircraft with minor output
seal leaks. Since the observed rate of ejection was relatively low compared to the flight duration on an actual aircraft, properly servicing the main gear compartment is sufficient for maintaining static mast grease levels.

2. Condition Indicator Improvements

A new thermal signature has been identified which can be related to lubricant service levels within a gearbox. A future health monitoring system could utilize measurements of thermal gradients rather than solely relying on over-temp limits to identify faulted conditions. Furthermore, the data continues to indicate that there are strong relationships between gearbox temperature and vibration signatures. From this information it may be possible to use temperature to normalize condition indicator levels. Enhanced, multi-sensor condition indicators could be more reliable and meaningful than the current vibration-based systems and lead to an enhanced application of Condition-Based Maintenance.

3. Temperature and Lubricant Issues

The thermal transients observed in the first two studies indicate the high possibility of thermal conditions within the AH-64 gearboxes, which exceed the operational limits of the lubricant, and that these events may occur without detection. Furthermore, changes in gearbox behavior observed during the third experiment indicate that the lubricant currently in use may be subject to rapid changes in physical properties, even at normal operating conditions. Follow-up research will be needed to further assess the operational characteristics of these gearboxes, and whether or not the grease has an appropriate specification for its current use.
4. Needed Improvements in the Dissemination of Information

The accepted practice of removing AH-64 tail rotor gearboxes for output seal leaks strongly illustrates a poor understanding of the operational characteristics of this component, and shortcomings in the dissemination of information between the manufacturer and customer. Anecdotes which have persisted during the life of the aircraft explained this maintenance practice by insisting that the static mast of the gearbox was sealed from the serviceable main compartment; however, also throughout the life of the aircraft there have been manufacturers and overhaul sites which would have been aware that such a seal never existed. Improved dissemination of facts between the Army, overhaul facilities, and the aircraft and component manufacturers would have prevented such costly misinformation from persisting for so long. In this particular case, the problem was confounded by the fact that the purported part was not accessible or visible to aircraft users. It is therefore recommended that maintenance practices based on unobservable components, which are also high cost drivers, be investigated to ensure that such policies are, in fact, needed and cost-effective.
CHAPTER 6

ANALYSIS OF LAB EXPERIMENT DATA: CONDITION INDICATORS

Introduction

HUMS Deployment on AH-64 Aircraft

Over the past ten years, the US Army has been actively installing and utilizing onboard Health and Usage Monitoring Systems (HUMS) for its fleet of AH-64 Apache helicopters. The primary purpose of this technology is to assist with rotor track and balance procedures and to perform condition monitoring of critical drive train components. Since its initial deployment, these onboard systems have made several thousand vibration measurements in a wide variety of locations, climates, and operating conditions which have been transferred and stored on a centralized data server. However, due to the complexity and volume of the data, traditional analysis methods have been unable to produce a significant number of major improvements to the reliability and usefulness of these systems. This study investigates improving fault diagnostic capabilities of onboard HUMS through the use of applied data mining.

Fault Identification Approaches

With the end user in mind, most HUMS packages assist aircraft maintainers in identifying faulted components through the use of simple interfaces and indicators. Since raw vibration signals are difficult to interpret and require large amounts of storage space, it is standard practice to transform sensor time series data into other domains or parameters through the use of a set of common functions specialized for a given part that is being monitored.

The most commonly utilized functions are called Condition Indicators (CI), which output a dimensioned or dimensionless single scalar value. These values are compared with established thresholds in a simple decision tree classifier which typically assign the
CI some form of ranked class such as “Good,” “Caution,” or “Exceeded,” and these classes are then utilized by maintainers in the maintenance decision-making process. A given component can have several Condition Indicators, and typically CIs are not fault-specific; multiple fault types can affect the value of a single CI, and a single fault could affect multiple CIs.

Some systems simplify the task of diagnostics by combining multiple CIs into a single Health Indicator (HI). This is done by weighted summation or some other combination method such that each component or sub-component has a single scaled value which expresses its overall health. Similar to CIs, HIs do not attempt to diagnose a specific fault type but rather to convey whether a part is in a good or deteriorating condition.

*Data Mining Driven Diagnostics*

The shortcomings of these methodologies are that not every CI has well-established thresholds, and HI combination functions may only be optimized to identify a particular fault type. Rather than attempt to improve these algorithms from a physical model or theory, a data mining approach can be utilized to find impending faults by observation. Faults can then be diagnosed through de facto relationships, regardless of their underlying causes. An illustration comparing this technique with the traditional approach is presented in Figure 6.1. The goal of this approach is to combine information from CIs in such a way as to identify specific faults within a component by using historical training data and case matching. The output of such a system would not provide the user with simplified generalizations such as good or bad, but rather it would output the most likely fault scenario for the article in question.
Although a large volume of CI data from actual aircraft exists, matching this data with component faults can be difficult, since HUMS data and maintenance records are not sufficiently integrated. This study examines a subset of the aircraft data to evaluate the usefulness of the currently utilized CIs and to determine whether the data contains sufficient resolution to distinguish faulted and non-faulted components. A data mining approach is then demonstrated on seeded-fault testing data generated in a laboratory setting to generate a fault classifier in order to test the feasibility of the concept. The classifier is then applied to other components from the same test stand as well as aircraft CIs to evaluate its performance.
Main Body

Aircraft Data Overview

The aircraft data used in this study was obtained from 54 AH-64A helicopters that were each equipped with a Vibration Monitoring Unit (VMU), which was the standard HUMS kit during the time of the data collection [26]. The entire set represents measurements taken over an eight year period which includes 2,100 distinct acquisitions and 261,000 CI values. A distribution of these acquisitions by tail number reveals that the aircraft and their respective HUMS devices had varying degrees of use (Figure 6.2).

![Figure 6.2 – Histogram illustrating the various amount of use of the onboard HUMS system by aircraft](image)

All data examined in this study was acquired at a flat-pitch ground at 101% rotor speed (FPG101) flight regime. For the given acquisition mode, there was a total of 138 CI functions which monitored 14 different components on the aircraft (Figure 6.3), and the statistical distribution of values was found to vary widely by CI. Through normalization, it is possible to qualitatively assess the relative usefulness of each CI by examining its statistical variance and the shape of its distribution. A CI with a wide
variance and flat distribution is more likely attributed to random probability or signal noise than to the actual condition of the component.

![Figure 6.3 – Sensor locations monitored by the Vibration Monitoring Unit [26]](image)

**Aircraft Dimensionality and Uniqueness**

Another aspect which should be examined for large datasets is the true dimensionality of the data, and whether or not the measured quantity represents a repeatable and meaningful metric. The issue of dimensionality can best be illustrated through the use of a cross-plot matrix, which shows the independence of each CI pair (Figure 6.4). Two CIs which are functionally dependent are essentially redundant, since the value of one can be predicted when given the value of the other. Utilizing a principle component analysis (PCA), it was found that 95% of the dataset variance could be represented by 84 orthogonal components, indicating that nearly 40% of the information contained in the original 138 CIs is redundant [2]. This dimensionality is still significantly complex, and averages to 6 measurable parameters per component being monitored.
The precision contained within the dataset can be demonstrated by building a classifier which attempts to predict individual aircraft tail numbers when only given the collected CI values. In theory, if the data contains enough information to distinguish between multiple healthy aircraft, then it is reasonable to assume that identifying failed components would be possible using similar techniques. A random forest classifier \(^\text{[27]}\) was built with 30 trees and 8 randomly selected CIs each, and a 10-fold cross validation was performed on the resulting model with the results displayed in the table below.

**Table 6.1 – Random Forest classifier results for predicting aircraft tail number from CI data**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly Classified Instances</td>
<td>1823 (85 %)</td>
</tr>
<tr>
<td>Incorrectly Classified Instances</td>
<td>320 (15 %)</td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.8472</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.0252</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.0996</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>70 %</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>74 %</td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>2143</td>
</tr>
</tbody>
</table>

These findings suggest that given the data from three distinct CI acquisitions from a given aircraft, it is possible to correctly identify its tail number 94% of the time. The uniqueness of each aircraft vibration signature is promising, since it implies the currently
deployed CI functions are sufficiently sensitive to profile the health of components; however, too much resolution may also lead to over-fitting of future developed classifier models. For example, if an aircraft has a known fault condition, a classifier may incorrectly characterize the fault by simply identifying the individual component, instead of identifying a more generalized parameter which indicates failure. Future applications may have to overcome this problem by adding additional noise to the data so that classifiers are optimally trained.

*Laboratory Test Data Overview*

The aircraft data available in this study was not supplemented with component fault histories, and therefore there was insufficient information to train classifiers to identify fault conditions. In order to demonstrate an applied data mining technique capable of characterizing a single fault type, it was convenient to utilize data generated from a laboratory setting instead. Prior studies on an AH-64 tail rotor drive train (TRDT) test stand (Figure 6.5) showed that when the output seal of a tail rotor gearbox is damaged, grease flows out of the main compartment through the static mast [4]. In those experiments, this phenomenon was demonstrated on three different gearboxes, all of which failed due to gear mesh lubricant starvation approximately 500 hours of operation after the initial seeding (Figure 6.6).
The three articles in the experiment had between 490 and 670 acquisitions, each of which contained 22 dedicated CIs to monitor the gearboxes as they transitioned to failure. In addition, 95 acquisitions were obtained from a gearbox with a different fault condition, 86 acquisitions from the third gearbox in this study prior to seeding, and 438 acquisitions from a non-faulted gearbox. To illustrate the sequence of experiments, a timeline of the TRDT test stand configuration is shown in Figure 6.7.

Figure 6.5 – AH-64 tail rotor drive train test stand

Figure 6.6 – Tail rotor gearbox teeth damage due to lubricant starvation experiments

Figure 6.7 – Timetable and record count for the test stand configurations which were examined. Articles 1-3 all failed due to gear mesh lubricant starvation
Gearbox Dimensionality and Uniqueness

Similar to the aircraft dataset, another PCA was performed on the 22 CIs, and found that 11 independent parameters were capable of describing 95% of the variance over the set of five gearboxes. When examining the same subset of indicators from the aircraft dataset, PCA found the dimensionality to be 14. This shows that data generated from actual aircraft has slightly more complexity than can be accounted for in the laboratory setting.

Also as before, a classifier was trained to determine if the gearbox CIs were capable of distinguishing between the individual gearboxes. For the test stand data, which consisted of only 5 possibilities, a random forest classifier was able to correctly identify the data source 98.3% of the time. For the aircraft case, in which there were 54 possibilities, tail rotor gearbox CIs from a single acquisition were sufficient to identify the aircraft tail number with an accuracy of 43.3%. It should be noted that contributing to this weaker classifier performance is the fact that during the eight years in which data was collected, the gearboxes or other neighboring components, such as the tail rotor swashplate, may have been replaced. Although relatively lower than the accuracies presented so far, it is remarkable to note that individual components generate signatures which are distinct enough to uniquely identify their aircraft on which they are installed.

Gearbox Fault Classifier Development

Because the three seeded fault gearboxes all failed within 580 ± 15% hours of operation, it was assumed that the same fault progression history and failure mechanism was achieved for all three articles. To accommodate the variation in the time-scales, each time parameter was divided by the total life of the article to generate a new parameter
defined as the gearbox life fraction. A cross-plot matrix was then utilized to determine which CIs changed as the articles approached failure. A total of 12 CIs were identified, including three tail rotor swashplate CIs, which also responded to the deteriorating condition of the gearboxes (Figure 6.8). Since no swashplate was actually installed during testing, these CIs, although interesting, cannot be applied to actual aircraft, and so classification was performed strictly on the tail rotor gearbox indicators instead.

It was also observed that Article 3 displayed the latest onset of condition change and had progressed the least by the end of its testing life. These observations are consistent with visual inspection of the gear teeth, which found Article 3 to be the least damaged at the end of the experiment [4]. A more detailed preprocessing of the data could be performed, which would compensate for these effects by adjusting the time scales accordingly; however, in order not to bias the results, no further modification was performed on the data.
The objective of this classification was to predict the life fraction of the gearbox when supplied with only the standard 22 tail rotor gearbox CIs. Originally, a k-nearest neighbors (KNN) classifier was attempted, since it is relatively simple to implement and its performance is well understood [2]. After normalizing the CI values on a scale from 0 to 1, a KNN classifier using five nearest neighbors was trained and tested using a 10-fold cross validation, with the results displayed in the table below.

Table 6.2 – Performance metrics of a KNN classifier which predicted gearbox life fraction from CI values

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient</td>
<td>0.8859</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.0941</td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.134</td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>37.6%</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>46.4%</td>
</tr>
<tr>
<td>Total number of instances</td>
<td>1688</td>
</tr>
</tbody>
</table>
In addition to the high correlation coefficient, a qualitative inspection of the classifier errors plot (Figure 6.9) shows that the incidence of severe false positives and false negatives is relatively low. Several other classifiers were attempted following these results; however, none of the more complex methods significantly outperformed the KNN method, and therefore those results are not presented.

![Classifier error plot comparing predicted gearbox life fraction to actual life fraction](image)

**Figure 6.9 – Classifier error plot comparing predicted gearbox life fraction to actual life fraction**

One concern in using cross-validation methods is the possibility of over-fitting, particularly since the number of test samples in this experiment was low. In order to test the classifier more stringently, data from Articles 1 and 2 were used to train separate classifiers, which were then applied to the opposite article. Using an identical KNN classifier scheme, data from Article 1 was found to be a poor predictor of Article 2 life fraction, but Article 2 was surprisingly good at predicting the life fraction of Article 1 (Figure 6.10). The corresponding classifier error plot further shows this one-sidedness.
A final step in classifier training is validation of the model with external data. Three different test sets were applied to the trained KNN classifier: data from un-faulted gearboxes on the test stand, data from a gearbox with a different fault, and data from the aircraft fleet which was examined previously. The classification results along with the mean predicted life fraction and standard deviation are shown in Figure 6.11.
Figure 6.11 – Predicted gearbox life fraction of healthy test stand gearboxes (top), a gearbox with a different type of fault (middle), and aircraft gearboxes (bottom)

From the first case, it can be seen that due to the relatively small training set, the classifier has been over-fit to the faulted gearboxes and therefore predicts higher life fraction values for the healthy gearboxes. This could be corrected by supplying the training set with additional healthy gearbox data, ideally consisting of a small number of samples from a large number of gearboxes [2]. The second case shows that the classifier may be capable of identifying other faults beyond what the training set had supplied. This may or may not be desirable, since for flight critical safety components, overall life is the only important factor. The final case, in which a test-data-derived classifier was applied
to actual aircraft, shows that limited data generated from a highly controlled environment is difficult to apply to systems which operate under much wider operating conditions. Other discrepancies may exist due to physically different hardware configurations, such as the laboratory test facility lacking a tail rotor swashplate assembly and tail rotor blades.

Conclusions

Data Mining Approach Possible

The current deployment of onboard HUMS devices on AH-64 helicopters generates sufficient data to produce data mining derived fault class identifiers. Although the present maintenance logging practices limit the short-term adoption of this approach, the condition indicator data already available shows a high degree of precision and will likely prove very effective when this technique ultimately implemented. Furthermore, experiments with laboratory test data shows that HUMS condition indicators are easily analyzed with the simplest of data mining algorithms. Future work may examine more advanced analysis, such as the mining of frequency domain data for spectrum-condition relationships.

Aircraft Uniqueness Detectible

One surprising finding in this study was that aircraft can be uniquely identified by their vibration signatures. This phenomenon could be observed using relatively easy-to-implement random forest classifiers on the lowest dimensioned data currently available: CIs. This ability may have no practical standalone application; however, this uniqueness of vibration signatures increases the likelihood that abnormal behavior or faulted components could be identified. It may be possible in the future to develop health
monitoring systems that do not diagnose a particular fault, but rather notify the maintainers whenever the condition is changing or abnormal.

*Remaining Life Prediction Capabilities*

By analyzing the collected set of condition indicators from laboratory test data, it was demonstrated that a classifier could be trained which could accurately predict the progression of a particular fault type in an AH-64 tail rotor gearbox. The successful prediction of faults which lead to the complete failure of a component shifts the topic of this research into the field of remaining life prediction and prognostics. Indeed, had the predicted life fraction metric been converted back into hours by multiplying by the average lifespan of the gearbox, this goal could easily have been reached. Without a successful application of the classifier to real aircraft, however, it would seem premature to assert that prognostics have truly been attained. Yet, these findings seem to suggest that the simplest solution to this goal may be through a data mining approach rather than the currently accepted practice of fatigue damage fraction summation.

*Future Improvements to CBM Practices*

The findings of this work illustrate some fundamental requirements for utilizing data mining to enhance the capability of component fault diagnostics as well as remaining life prediction. Comprehensive efforts such as the Army Condition-Based Maintenance (CBM) program could benefit from such strategies only if it maintains readily accessible component failure histories which are reasonably linked to the vibration data. To date, the ability to strongly relate HUMS information with actual faults has been limited to a select number of cases which have been used to highlight the abilities of the onboard systems. Program improvements should seek to link teardown
analysis results to flight histories and vibration data to begin the necessary process of growing an adequate training data set to enable this new approach.

**Acknowledgements**

This research was performed in support of the continued improvement of Condition-Based Maintenance efforts by the South Carolina Army National Guard and the Army Engineering Directorate. The authors would like to thank all parties involved in assistance with data acquisition and technical support.
CHAPTER 7
ANALYSIS OF LAB EXPERIMENT DATA: SPECTRA

Overview

This chapter will extend upon the experiment presented in Chapter 6 by examining the spectral data that was collected from the gearboxes. Although raw time series data had been collected for these articles, it was decided that intermediate-level data, i.e., the vibration amplitude spectra, provides more similarity with data that is generally available from aircraft and other fielded systems. The objective of this section is to build classification models which can distinguish various fault severities by examining the provided spectra. Due to the very large dimensionality of each measurement instance, appropriate techniques of dimension reduction will be deployed. Also, it is assumed that by utilizing the higher resolution dataset, issues such as overfitting will become an even greater problem. Therefore, this section will examine the classifier model accuracy with both cross-validation and designated test sets.

Characteristics of Data Set

For the three different specimens in this study, two accelerometers were mounted perpendicular to each other: the first was oriented radially with respect to the input gear and the second was mounted on the gearbox housing approximately along the input gear axis of rotation. Relative to the mounting orientation of the gearbox, the radial and axial accelerometers are termed lateral and vertically mounted, respectively. Measurements
were collected approximately once per hour over the several hundred hours each gearbox survived at a sampling rate of 48 kHz and a sampling window of 0.3413 seconds, or 16384 samples. These produced vibration amplitude spectra over the frequency range of 0 to 24 kHz divided into 8192 frequency bins, giving a frequency resolution of 2.93 Hz. Since there are two common units of measurement for gearboxes in this range of running speeds, both the raw acceleration and calculated velocity spectra were examined initially. Each gearbox was run in a predetermined loading cycle which was described previously (Table 5.1), and detailed information on the total hours of life and number of measurement samples was given previously in Figure 6.7.

**Pre-filtering Techniques**

As a result of the stepped loadings applied to each gearbox, the observed temperatures also followed a similar stepped pattern. It was also observed that the vibration readings were influenced somewhat by the temperature at the time of measurement, possibly due to changes in lubrication properties, material stiffness, or thermal effects on the accelerometer. Since it is the purpose of this study to build classifiers which relate mechanical condition to the vibration spectra, it was decided to reduce these observed fluctuations by performing a 5-point moving average filter to each frequency bin.

**General Observations**

Examining the combined spectrographs of the three articles (Figure 7.1), shows somewhat similar behavior for each on both sensor locations and units of measure, in that all graphs show increased amplitudes for frequencies greater than 8 kHz as the article approach their final failure condition. There are also a number of significant differences
which can be observed for each article. Article 1 had the shortest life and showed an abrupt condition change which appears to diminish until it experienced an over-limit temperature failure. Article 2 shows a consistent increase in signal magnitudes throughout its life until it failed when an input gear tooth broke. Article 3 had the longest life, and shows some abnormally high signal levels in the 2 to 8 kHz band even when the article was assumed to be in a healthy state. It showed the least change as it approached failure, but it is also noted that the end-state condition of this article was not as severe as the other two, and that termination was defined by a more conservative temperature limiting technique.

Figure 7.1 – Complete filtered dataset for acceleration (top) and velocity (bottom) spectra for the vertical (left) and lateral (right) sensor locations

87
RMS Values

In order to simplify the interpretation of the accelerometer spectrographs, the RMS of each spectrum was computed and plotted (Figure 7.2). It can be seen that both sensors show similar trends in the overall acceleration signal level, however, these results are largely dissimilar from the RMS plots of the velocity spectra. This is likely because, as observed earlier, the largest changes in the vibration spectra as the components approached failure occurred in the higher frequencies, which are attenuated by the conversion from acceleration to velocity. As a result, there is less of a pronounced change in signal level for that unit of measurement. Another interesting observation from examining the total signal magnitude is the large difference in paths towards failure. Of note is the sudden jump which occurs in the middle of the life of Article 1 and at the end of the life of Article 3, which is replaced by a seemingly more gradual trend in Article 2.

The major conclusion to draw from examining these RMS plots, as well as the spectrograms, is that even in a controlled laboratory setting, the same component with the same fault condition may fail in very different ways. Also, since the RMS of the velocity spectra are less fitting to the ideal monotonically increasing failure indicator, for simplicity only the acceleration spectra are to be used in training fault classifiers.
Figure 7.2 – RMS signal values for acceleration (top) and velocity (bottom) spectra for the vertical (left) and lateral (right) sensor locations

**Compressibility**

Prior to supplying the vibration spectra to classifier training, it was noted that significant dimension reduction is both needed and easily achieved. Extending the techniques of Chapter 3, singular value decomposition (SVD) was performed on all four measurement sets and the normalized diagonals of the singular value matrix are plotted below. Shown on a log-log plot, the singular values diminish linearly until limitations in sample size cause a much sharper decrease in information representation. This linear log-log trend appears throughout this work to be a universal property of vibration spectra matrices.
As noted in previous chapters, these normalized singular values are equal in magnitude to the relative RMS error incurred when reducing to a specified dimensionality. For both sensors and both units of measure, projecting the 8192 bin data to 100 attributes would result in less than a 1% error after decompression. Therefore, for this study, all SVD reduction chooses 150 singular vectors per sensor, resulting in a 98% reduction in model complexity with less than 1% information loss.

![Error fraction curves for the studied units of measure and sensor locations](image)

**Figure 7.3 – Error fraction curves for the studied units of measure and sensor locations**

*Approximating Fault Class*

In order to supply a training set to a classifier, some estimation of fault condition has to be established. Although it has been noted that each article exhibited differing
paths to failure, a simple exponential failure was estimated for each article. In contrast to the previous chapter which assumed a linear progression towards failure, this approach allowed for custom rate parameters to be selected such that the approximate fault metric followed the observed signal RMS trends (Figure 7.4). The estimated fault parameter was then discretized into four equal magnitude classes, which are labeled as Good, Abnormal, Caution, and Failed.

![Figure 7.4 – Approximated damage fraction used for classification training](image)

**Classification of Source Article**

Since it was observed that a relatively small number of condition metrics are capable of uniquely identifying the source article, it was assumed that similar characteristics would be observed for the more complex vibration spectra data. This could lead to the problem of overfitting, in which the given model correctly identifies the fault class of an article by taking into account its source rather than approximating a general model for all articles of the same type. This is especially true in situations where
the number of unique data sources is limited, such as laboratory data which produces a large number of samples from a relatively small number of articles.

**Distance Matrix Characteristics**

One approach to understanding the susceptibility a given dataset has to overfitting is to examine the pair-wise distance between each measurement. Since all the features of a given spectrum instance represent component magnitudes and have the same units, it is possible to define the distance between to spectra as:

$$ D(X_1, X_2) = \sqrt{\sum_n (x_{1,n} - x_{2,n})^2} $$

where $x_{k,n} \in X_k$. This definition is equivalent to the RMS of the difference between the two spectra. It should be noted that $D(X_1, X_2) = D(X_2, X_1)$, and therefore the distance matrix $D_{m,n} = D(X_m, X_n)$ for a given set of measurements is symmetric. As a result, it is possible to show both the lateral and vertical sensor measurement distances in a single combined plot by normalizing each sensor distance matrix by its maximum value and combining the upper and lower triangles (Figure 7.5). A similar distance matrix was computed for the velocity spectra (Figure 7.6).

From the acceleration distance matrix plot, the first observation that can be made is the apparent self-similarity each article has with itself prior to any major advancement of the fault condition. This implies relatively good repeatability of the measurements, and for Articles 1 and 2, there are also additional regions of self similarity along the main diagonal which indicate that the transition to failure consists of relatively large steps followed by periods of relative stability. Review of the velocity distance matrix is less notable, with the exception of some previously unobserved transients visible as thin lines.
on the lateral sensor, and an otherwise hidden bi-state experienced by Article 3 observed by the vertical sensor. The strong self-similarity observed in the acceleration spectra and the relatively larger distance between healthy measurements from different articles implies that by including the full information contained in the vibration spectra, distinguishing the source of the measurement would be relatively successful, while the classification of fault class may be less so. The challenge then becomes to remove the information which uniquely identifies a given gearbox, and leave behind data which relates to fault condition only.

Figure 7.5 – Normalized Euclidian distance matrix of the acceleration spectra for the lateral (upper left) and vertical (lower right) accelerometer locations
Figure 7.6– Normalized Euclidian distance matrix of the velocity spectra for the lateral (upper left) and vertical (lower right) accelerometer locations

Classifier Results

Noting the similarity between the lateral and vertical acceleration measurements, an SVD reduction was performed on the combined 16384 attribute dataset to produce a single 300 element singular vector per measurement. These measurements were then tagged with the source article and supplied to four different classifier methods. For each method, a 10-fold cross validation was performed and the combined results are presented below (Table 7.1).

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Correctly Classified</th>
<th>Incorrectly Classified</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN</td>
<td>1581</td>
<td>100</td>
<td>94.1%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>1418</td>
<td>263</td>
<td>84.4%</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>1677</td>
<td>4</td>
<td>99.7%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>1657</td>
<td>24</td>
<td>98.6%</td>
</tr>
</tbody>
</table>
These findings are consistent with the observed distance matrix characteristics and indicate that any attempts to produce classification models from this dataset should include methods other than cross validation to evaluate its performance.

Classification of Fault Class

Cross-Fold Validation

Utilizing the reduced 300 feature dataset from all three test articles, another set of classifiers were built which attempted to predict the guessed fault class of each measurement. The classifiers were evaluated, once again, using 10-fold cross validation and the results are presented below (Table 7.2). Since the data involved was not a simple binary class, the performance metrics of true positive (TP), false negative (FN), and false positive (FP) refer to the overall rate of correctly classified instances, the rate at which a measurement was classified into a less severe fault class, and the rate at which an instance was classified into a more severe fault class, respectively.

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>Actual Class</th>
<th>Class TP Rate</th>
<th>Overall Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>kNN</td>
<td>G</td>
<td>1325</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>2</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Naive Bayesian</td>
<td>G</td>
<td>1148</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>143</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>G</td>
<td>1307</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>15</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Random Forest</td>
<td>G</td>
<td>1314</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>8</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
From this table, it can be seen that the first 300 elements in the spectra singular vectors is capable of accurately identifying a particular fault class. This initial proof-of-concept classification results are able to distinguish various states of failure without any intrinsic knowledge of the system being monitored or any conventional condition metrics. There is also no clear distinction on classifier performance; however, it should be noted that interestingly these results are very similar to those obtained when attempting to classify for the source article. This may be indicative that all the tested classifier models rely on training data from the given article in order to make an accurate class prediction.

In order to get a better assessment of these classification techniques, three additional experiments were performed in which one article was designated as the test set, while the other two were used as the training set. Initially this technique was found to produce severely poor results, with nearly all of the predicted results being the Good class. It was concluded that this was the result of class imbalance in the training set, since there are in all cases a much larger set of Good samples than any of the other three classes. As a result, each training set was re-sampled to 70% of its original size, thus producing a more uniform class distribution in given training sets. The results of the one-out classification tests are shown in Table 7.3, Table 7.4, and Table 7.5, below.
Table 7.3 – Classifier performance results trained on Articles 2 and 3 and tested on Article 1

<table>
<thead>
<tr>
<th>Classifier Method</th>
<th>Predicted Class</th>
<th>Actual Class</th>
<th>Overall Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>kNN</td>
<td>238</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>Naive Bayesian</td>
<td>231</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>238</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Random Forest</td>
<td>238</td>
<td>13</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 7.4 – Classifier performance results trained on Articles 1 and 3 and tested on Article 2

<table>
<thead>
<tr>
<th>Classifier Method</th>
<th>Predicted Class</th>
<th>Actual Class</th>
<th>Overall Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>kNN</td>
<td>374</td>
<td>77</td>
<td>46</td>
</tr>
<tr>
<td>Naive Bayesian</td>
<td>297</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>317</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Random Forest</td>
<td>310</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Classifier Method</td>
<td>Predicted Class</td>
<td>Actual Class</td>
<td>Overall Performance</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>kNN</td>
<td>626</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>Naive Bayesian</td>
<td>625</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>626</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Random Forest</td>
<td>625</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

It can be seen immediately that the performance of these classifiers is significantly less than those trained on cross validation. It is also noteworthy that in this scheme, the four classification methods appear to perform more similarly, and the naïve Bayesian method no longer produces the lowest accuracy. In the case Article 3, the results are seemingly better than the other two; however, close examination of the distance matrix finds this is due to class imbalance in the test set. For example, the kNN classifier predicted a Good class label 100% of the time, but still achieved an overall accuracy of 92.3% since was the most common fault class in the test set. These results cast doubt on whether there is a sufficient sample size in the training set to correctly produce underlying fault detection models, particularly with regards to the number of distinct articles rather than the number of measurements per article.
In Chapter 8, it was shown that utilizing reduced forms of vibration spectra for fault classification produce similar results to classification methods which are based upon engineered condition indicator metrics. The shortcomings of the classifiers in either case were found to be the small number of test samples on which to establish system to system variation, which is often large relative to the changes in signal observed as an article transitions to failure. This chapter extends the techniques of fault classification on vibration amplitude spectra by examining a diverse sample set of bearings and gearboxes and comparing the findings with a set of observed gearbox failures.

Baseline Dataset

The dataset used in establishing the baseline characteristics is an extended version of that presented in Chapter 3. In addition to the three gearbox types, three types of bearings are added to this set to provide a more generalized view of the implications of examining very large field datasets. It should be noted that all these mechanical components were observed on a single type of rotorcraft, and in general, the measurements were taken concurrently from all sensor types as required by the preventative maintenance program for that particular platform. Drive train parameters such as speed and power level are theoretically consistent between measurements;
however, these were not measured and thus cannot be verified. Specific details about the
gearbox and bearing types are given in Table 8.1 and Table 8.2, respectively. For both
gearbox Type I and bearing Type I, there are two different positions where the article can
be installed, and these sample sets are separated by the designation A and B. No pre-
filtering was performed, and the complete set of spectrograms is shown in Figure 8.1.

Table 8.1 – Detailed specifications of gearboxes examined in this study

<table>
<thead>
<tr>
<th></th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor Type</td>
<td>Dytran 3077A</td>
<td>Dytran 3062A</td>
<td>Dytran 3077A</td>
</tr>
<tr>
<td>Mounting Style</td>
<td>Bracket</td>
<td>Stud</td>
<td>Bracket</td>
</tr>
<tr>
<td>Mounting Direction (relative to input)</td>
<td>Axial</td>
<td>Radial</td>
<td>Radial</td>
</tr>
<tr>
<td>Gear Mesh Frequency (Hz)</td>
<td>12910</td>
<td>2999</td>
<td>1346</td>
</tr>
<tr>
<td>Sampling Rate (kHz)</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Samples per Acquisition</td>
<td>8192</td>
<td>8192</td>
<td>8192</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Number of Specimens</td>
<td>89</td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td>Number of Acquisitions</td>
<td>960</td>
<td>926</td>
<td>1041</td>
</tr>
<tr>
<td>Average acquisitions per specimen</td>
<td>10.9</td>
<td>10.6</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Table 8.2 – Detailed specifications of the bearings examined in this study

<table>
<thead>
<tr>
<th></th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor Type</td>
<td>Dytran 3077A</td>
<td>Dytran 3077A</td>
<td>Dytran 3077A</td>
</tr>
<tr>
<td>Mounting Style</td>
<td>Bracket</td>
<td>Bracket</td>
<td>Bracket</td>
</tr>
<tr>
<td>Mounting Direction</td>
<td>Radial</td>
<td>Axial</td>
<td>Axial</td>
</tr>
<tr>
<td>Outer Race Pass Frequency (Hz)</td>
<td>288</td>
<td>196</td>
<td>423</td>
</tr>
<tr>
<td>Sampling Rate (kHz)</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Samples per Acquisition</td>
<td>4096</td>
<td>2048</td>
<td>2048</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Number of Specimens</td>
<td>89</td>
<td>89</td>
<td>89</td>
</tr>
<tr>
<td>Number of Acquisitions</td>
<td>965</td>
<td>966</td>
<td>896</td>
</tr>
<tr>
<td>Average acquisitions per specimen</td>
<td>10.8</td>
<td>10.9</td>
<td>10.1</td>
</tr>
</tbody>
</table>
Compressibility

A similar investigation of singular value decomposition was performed on each of the eight baseline data sets and found similar results to those discussed in previous chapters. For consistency, all dimension reduction utilizes SVD to project the varying attribute features of 2048, 4096, or 8192 down to the first 300 elements of the right singular vector for each data matrix.
Figure 8.1 – Full spectrograms of gearbox (left) and bearing (right) datasets examined
Distance Matrix Characteristics

Once again, prior to attempting any form of classification from the vibration spectra, it is useful to understand the sources of measurement variation. Utilizing the Euclidian distance matrix allows for the comparison of spectral changes between two measurements on the same specimen, as well as from article to article. Sources of variance in the measurement set include manufacturing variations, differences in loading characteristics, environmental effects, sensor repeatability, and also mechanical faults. A distance matrix cannot distinguish between these sources of variation, but does provide insight on the relative impact of each of these sources may have.

Utilizing only the acceleration amplitude spectra, a complete distance matrix was computed for each article type and plotted (Figure 8.2). The sample index was sorted by specimen and date such that multiple measurements from the same article appear consecutively near the matrix diagonal; however, the sample numbers are generally not related between gearbox and bearing types. From these graphs, a number of interesting observations can be made. First, as expected, the lowest distances appear near the diagonal and imply high repeatability when measurements are taken from a single specimen. Second, the presence of lines which span the entire matrix indicate that the general population as a defined central tendency, and that an increased distance from one article generally implies an increased distance from the entire population, i.e., in general, there appears the measurement distributions are not multimodal. Finally, at several locations, it can be observed that pairs of specimens that are distant from the main group are sometimes relatively close to each other, implying the cause of this variation is consistent between multiple articles.
Figure 8.2 – Euclidian distance matrices of the vibration amplitude spectra for the four gearbox (left) and bearing (right) datasets
Figure 8.3 – Probability density distribution of the Euclidian distance between multiple measurements from the same article (red) and measurements from different articles (blue)
The next step is to quantitatively assess the difference in variation between measurements and between articles. For this, each distance matrix was divided into two samples, one representing measurements between different specimens, and the other between measurements from the same specimen. A 100 bin histogram was then created and the results were converted into probability density curves for each article type and measurement type (Figure 8.3). From this graph, it is clear that the variation from specimen to specimen is significantly larger than that experienced from different measurements on the same article.

A receiver operator characteristic (ROC) curve was created to establish the level of accuracy at which pair-wise distance between two spectra can predict whether two given measurements originate from the same specimen. The areas under the curve are presented in Table 8.3, and indicate, for example, that for the case of Type I bearings, distinguishing between two source articles and two measurements from the same article can be achieved more than 86% of the time.

Table 8.3 – Receiver Operator Characteristic (ROC) area for the Euclidian distance metric

<table>
<thead>
<tr>
<th>Article Type</th>
<th>Bearings</th>
<th>Gearboxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>I A</td>
<td>0.863</td>
<td>0.670</td>
</tr>
<tr>
<td>I B</td>
<td>0.900</td>
<td>0.703</td>
</tr>
<tr>
<td>II</td>
<td>0.810</td>
<td>0.734</td>
</tr>
<tr>
<td>III</td>
<td>0.724</td>
<td>0.765</td>
</tr>
</tbody>
</table>

Classifier Performance

As before, a set of classifiers were trained on an SVD-reduced 300 feature dataset to predict the originating article and were evaluated with a 10-fold cross validation (Table 8.4 and Table 8.5). The results show that most classifiers were still able to distinguish their specimen even though the number of distinct articles was greater than 87 for all cases and the number of measurements per was generally less than 12. This indicates that
the potential for fault classifier overfitting is not only the result of too few representative samples, such as shown in the previous Chapters, but also because the number of features selected, in this case, 300, may be too large to sufficiently generalize to sample space. Additional steps are needed in order to eliminate those features which have the greatest contribution to source article determination while preserving those which have a strong ability to separate faulted from normal measurements. Another method for handling the uniqueness of data from each specimen would be to derive article-specific thresholds or differential trending algorithms to diagnose large changes in the spectra rather than universally defined fault characteristics.

### Table 8.4 – Source gearbox classifier performance results

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Type IA</th>
<th>Type IB</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN</td>
<td>59.90%</td>
<td>61.88%</td>
<td>72.91%</td>
<td>71.85%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>77.08%</td>
<td>77.32%</td>
<td>75.98%</td>
<td>77.91%</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>50.10%</td>
<td>48.81%</td>
<td>44.28%</td>
<td>47.17%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>60.52%</td>
<td>61.88%</td>
<td>54.27%</td>
<td>62.92%</td>
</tr>
</tbody>
</table>

### Table 8.5 – Source bearing classifier performance results

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Type IA</th>
<th>Type IB</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN</td>
<td>63.52%</td>
<td>53.62%</td>
<td>48.66%</td>
<td>18.96%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>83.63%</td>
<td>84.99%</td>
<td>80.25%</td>
<td>64.13%</td>
</tr>
<tr>
<td>C4 Tree</td>
<td>61.45%</td>
<td>63.56%</td>
<td>58.82%</td>
<td>32.92%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>71.92%</td>
<td>72.67%</td>
<td>58.59%</td>
<td>33.37%</td>
</tr>
</tbody>
</table>

### Faulted Dataset

In previous chapters, attempts to classify component condition from either engineering-derived condition indicators or directly from the vibration spectra faced the problem of source-specimen overfitting. More specifically, although cross validation results were relatively successful, the predictive modeling capabilities were limited when there were no representative samples of a particular specimen in the training set. It was
hypothesized that this was due to an insufficient number of unique specimens, and so when examining the same approach for fleet data, it is desirable to have a larger number of unique specimens.

In previous work performed by Branning et al [28], a number of rotorcraft with a particular gearbox fault type were identified, and there exists data from these aircraft before and after the gearboxes were eventually replaced. Expanding upon that work, eleven of those aircraft were selected to attempt vibration spectra classification. In the spectrogram below (Figure 8.4), individual articles are separated by magenta dividing lines, and the measurement data is represented by the overlaid black lines. Vertical dashed lines denote when a given gearbox was changed on a given aircraft. It can be observed there is a large amount of variation in the number of samples for each article.

Figure 8.4 – Complete gearbox dataset with tagged failed and un-failed articles
Definition of Fault Class

Unlike the previous laboratory study, no attempt was made to produce a multi-class fault scheme for the training set. Since relatively little is known about the gearboxes from this dataset, each measurement was tagged as either Good or Failed based on the simple rules that any measurement that took place prior to replacement represents a failed article, and that newly installed gearboxes were all good.

Attribute Selection

Since the behavior of over-fitting in the form of article discernment has been demonstrated multiple times, a similar attempt to simply identify which gearbox each measurement represents was not performed initially. Rather a new approach was attempted to remedy the problem in which the information gain of each attribute was determined with respect to each given class label. Attributes which showed a high information gain against the fault class and a low gain for the source article class were prioritized and a shortened list of these features is shown below (Table 8.6).

Table 8.6 – Information gain metrics for selected reduced dimension features for fault-tagged dataset

<table>
<thead>
<tr>
<th>Feature Number</th>
<th>Fault Class Information Gain</th>
<th>Source Article Information Gain</th>
<th>Fault / Source Gain Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>0.1915</td>
<td>0.1271</td>
<td>1.507</td>
</tr>
<tr>
<td>41</td>
<td>0.3229</td>
<td>0.3059</td>
<td>1.056</td>
</tr>
<tr>
<td>22</td>
<td>0.3313</td>
<td>0.3244</td>
<td>1.021</td>
</tr>
<tr>
<td>88</td>
<td>0.2127</td>
<td>0.2084</td>
<td>1.020</td>
</tr>
<tr>
<td>36</td>
<td>0.1847</td>
<td>0.2303</td>
<td>0.802</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Fault Classification

From this sorted list, the top 4 features were selected to be supplied to the standard set of classifier methods. Plotting these features in a cross-plot matrix shows high separability between the Good and Failed class labels (Figure 8.5).

![Crossplot matrix showing separability of failed and unfailed gearboxes](image)

*Figure 8.5 – Crossplot matrix showing separability of failed and unfailed gearboxes*

The initial assessment was performed using 10-fold cross validation, and additional classification models were built and tested by excluding articles labeled 5195, 5508, and 5363 from the training set and using them as the designated test set. These particular articles were chosen due to their relatively large number of good and bad class data points. Classifiers were then trained on data with the entire set of attributes as well as the reduced set of four attributes and the results are shown below (Table 8.7).
Table 8.7 – Fault class classification results with 200 features vs. the optimized 4 features

<table>
<thead>
<tr>
<th>Test Article or Method</th>
<th>Cross Validation</th>
<th>5195</th>
<th>5508</th>
<th>5363</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>98.9%</td>
<td>31.6%</td>
<td>55.0%</td>
<td>35.2%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>91.6%</td>
<td>36.8%</td>
<td>64.2%</td>
<td>85.6%</td>
</tr>
<tr>
<td>C4</td>
<td>94.0%</td>
<td>36.8%</td>
<td>41.5%</td>
<td>31.5%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>98.7%</td>
<td>35.1%</td>
<td>58.2%</td>
<td>33.2%</td>
</tr>
<tr>
<td>Selected Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>83.3%</td>
<td>59.6%</td>
<td>84.2%</td>
<td>85.8%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>79.0%</td>
<td>76.3%</td>
<td>88.0%</td>
<td>94.5%</td>
</tr>
<tr>
<td>C4</td>
<td>83.8%</td>
<td>73.7%</td>
<td>79.4%</td>
<td>90.2%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>88.4%</td>
<td>61.4%</td>
<td>80.5%</td>
<td>88.7%</td>
</tr>
</tbody>
</table>

Source Gearbox Classification

Table 8.8 – Fault class classification results with 200 features vs. the optimized 4 features

<table>
<thead>
<tr>
<th>Test Article or Method</th>
<th>Cross Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Features</td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>98.6%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>96.8%</td>
</tr>
<tr>
<td>C4</td>
<td>94.2%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>97.7%</td>
</tr>
<tr>
<td>Selected Features</td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>59.3%</td>
</tr>
<tr>
<td>Naïve Bayesian</td>
<td>50.3%</td>
</tr>
<tr>
<td>C4</td>
<td>63.5%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>72.3%</td>
</tr>
</tbody>
</table>
Discussion

Implications of Specimen to Specimen Variance

Throughout this research the seemingly universal phenomenon of being able to identify the source of a given measurement has shown that, in general, the variation from measurement to measurement is small compared with the variation due to differences in multiple articles of the same type. This was observed in cases when there were a low number of samples and a large number of specimens, as well as the opposite case of a large number of samples and few specimens. Classifier performance was high for both gearboxes and bearings over varied and diverse operating conditions. It should be noted that since nearly all the data used in the study originated from rotorcraft components with presumably tight quality control requirements during fabrication, it is likely that this uniqueness property does not reflect poor manufacturing, but possibly a universal characteristic of drive train components.

Specificity and the Number of Features

Considering the abstract multi-dimensional feature domain, the relatively small distances represented in the entire space are not useful in determining component condition. Extracting the useful information for fault classification requires careful projection onto lower dimensional spaces which lose the information specific to the
individual source article. The previous chapter in this work demonstrated that this is possible, and that such projections need not be complex, and may be as simple as rotational transformation matrices. The incredibly small number of remaining features, in this case 4 from an original set of 200 features extracted from an 8192 dimensional vector, casts doubt on whether these simple reduction and selection techniques will be capable of producing advanced multi-class and multi-fault diagnostics.

The effects of overfitting are not just limited to unique source article identification, and also manifest as differing goodness of fit between training and test sets even when both sets are randomly selected from the same article at the same condition state. This implies that the generalized model of vibration signature for a given component may be very simple indeed, and that a generalized description of any vibration spectrum can likely be done with fewer than 100 features. Beyond that, the reduction scheme or classifier begins to fit the model to individual instances which are not uniformly found within the entire population dataset.

**Conclusions**

*Information Content and Data Requirements*

This research has consistently shown that vibration information, in the form of amplitude spectra as well as sets of condition metrics, tend to be highly redundant and overly-precise for the application. The true amount of information contained within a given measurement varies depending on running speed of the machine and the sampling frequency of the original time domain signal, but for all data studied in this research, compression ratios of 6:1 up to 50:1 seem reasonable. This is consistent with traditional
audio compression capabilities, and reflects that, in general, mechanical vibrations produce fairly simple time domain waveforms.

For systems running under fairly constant operating conditions, it was shown that the central tendencies of a given fleet of fielded components can be represented within a reasonable number of records and file storage requirements. It is expected that due to continuously decreasing costs of digital storage and transmission bandwidth, this aspect will become even less of an issue. The combination of low information content and inexpensive storage makes possible the large scale archival and analysis of vibration signatures, providing new insights into the variability present within field measurements and new abilities to create enhanced diagnostics based on this information.

Data Mining Approaches are Successful

Apart from some problems with model overfitting, data mining approaches have been demonstrated to be an effective alternative to traditional physical parameter diagnostic metrics. Methods for dimensionality reduction were shown to be incredibly effective at preserving information content while reducing model complexity, and consistent relationships between the number of selected features and the expected compression loss were shown. Furthermore, utilizing attribute selection algorithms, the reduced dimension set could be further tailored to reflect a desired parameter, in this case fault class membership. The final step was in performing the classification, which in the case of cross-fold validation produced highly accurate descriptive models, and in the case of single-specimen-removed produced predictive models of varying accuracy. It is believed that further enhancements to these techniques would be capable of producing
fully automated diagnostic agents which preclude the need for any physics-derived metrics.

*Shortcomings with Traditional Condition Monitoring Approaches*

As a result of the common overfitting observation, there appears that the traditional model for condition monitoring may be insufficient to produce high fidelity results on diverse field applications. The observed uniqueness between specimens implies that regardless of the diagnostic algorithm used, each article may have its own definition of fault class. Utilizing a single set of indicators with a single set of thresholds is inherently problematic, since it has been found that a healthy component can produce signatures equivalent to a faulted component. It would appear that a more important aspect of diagnostics is state change or the technique known as trending. This could be achieved by monitoring rate changes in engineered diagnostic metrics, or by utilizing adaptive classifiers on each component which compare a given measurement to several of the previous measurements or perhaps to the initial measurement state.

*Recommendations*

*Wider Class of Operating Parameters*

Nearly all the data studied in this work are obtained from rotorcraft under a scheduled preventative maintenance program. As a result, the operating parameters of running speed and torque were generally consistent from measurement to measurement. Variations in torque or running speed would significantly change the vibration signature and it is theorized that data mining algorithms would be able to normalize for these input parameters and produced diagnostics regardless of operating condition.
*Complex Valued Spectra*

Literature review and early trials of complex valued spectra show that the dimensionality reduction techniques are equally as effective when phase information has been preserved. This opens up new opportunities to compress vibration signals in such a way that the original waveform can be approximated after decompression and inverse transform. In theory, this might allow time-domain based indicators to be extracted from the highly compressed spectra and other common techniques such as synchronous time averaging to be used.

*Expanded Methods of Feature Definition and Selection*

The currently completed work examines linear transformations, either in the form of principal components analysis or singular value decomposition in order to define the underlying information features. Although these feature spaces seem to be relatively small and good fits for the given data, this observed characteristic may be due to localized linearity. A more general approach may be to consider higher order polynomial expressions for the domain transformation which may result in a more useful set of feature definitions. Also, selection of features which are uniformly distributed among the sample training population is key to overfitting avoidance. It is believed that the remarkably small number of features remaining after this stringent selection process is likely too small to produce the desired detailed fault classifiers. Future work should focus on feature selection criteria which produce general models, but still have enough complexity to convey multiple instance states.
Specimen-Specific Classifier Models

A final suggested area for future work takes the reverse position of the objectives of this study. Due to the overwhelmingly self-similar nature of measurements between multiple acquisitions of the same specimen, it may be more effective to use this property to redefine traditional condition monitoring models to a more specimen-specific approach. Utilizing the same data mining techniques, a simplistic anomaly detection model which compares an observed state against historical observations for that model alone could be trained. For serviceable or repairable components, this could lead to adaptive fault classifiers as a maintainer logs activities performed and builds up a sufficient article history. This type of article-specific diagnostic metric could be used as a standalone state change detector or in conjunction with generic universal classifiers.

Final Remarks

The results of this study present a strong case for data-driven research methods in the field of condition monitoring. Although there are many caveats when performing this type of analysis, it is likely that the application of data mining and machine learning techniques will eventually outperform the traditional human engineered metrics. Continued expansion into this field will likely lead to improvements not only in diagnostics, but also create robust and diverse prognostic algorithms for future systems. Long term, data-driven research may ultimately have impacts beyond the maintenance of existing systems, but will also spread into all aspects of machine life cycle management, including design and manufacturing processes. The next generation of smart-machines, with the ability to observe their own strengths and weaknesses, will be able to adapt various aspects of their own design to sustain a longer, more useful product life.
REFERENCES


