

Anomaly Detection in Rolling Element Bearings via Two-Dimensional Symbolic Aggregate Approximation

Bradley W. Harris

Thesis submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Mechanical Engineering

Michael J. Roan, Chair
Robert L. West
Alexander Leonessa

May 2, 2013
Blacksburg, Virginia

Keywords: Symbolic dynamics, Rolling element bearings, Anomaly detection, Predicative maintenance, Digital signal processing, Bayesian statistics

Anomaly Detection in Rolling Element Bearings via Two-Dimensional Symbolic Aggregate Approximation

Bradley W. Harris

ABSTRACT

Symbolic dynamics is a current interest in the area of anomaly detection, especially in mechanical systems. Symbolic dynamics reduces the overall dimensionality of system responses while maintaining a high level of robustness to noise. Rolling element bearings are particularly common mechanical components where anomaly detection is of high importance. Harsh operating conditions and manufacturing imperfections increase vibration innately reducing component life and increasing downtime and costly repairs. This thesis presents a novel way to detect bearing vibrational anomalies through Symbolic Aggregate Approximation (SAX) in the two-dimensional time-frequency domain. SAX reduces computational requirements by partitioning high-dimensional sensor data into discrete states. This analysis specifically suits bearing vibration data in the time-frequency domain, as the distribution of data does not greatly change between normal and faulty conditions.

Under ground truth synthetically-generated experiments, two-dimensional SAX in conjunction with Markov model feature extraction is successful in detecting anomalies ($> 99\%$) using short time spans (< 0.1 seconds) of data in the time-frequency domain with low false alarms ($< 8\%$). Analysis of real-world datasets validates the performance over the commonly used one-dimensional symbolic analysis by detecting 100% of experimental anomalous vibration with 0 false alarms in all fault types using less than 1 second of data for the basis of ‘normality’. Two-dimensional SAX also demonstrates the ability to detect anomalies in predicative monitoring environments earlier than previous methods, even in low Signal-to-Noise ratios.

ACKNOWLEDGMENTS

I was very fortunate to work under the direction of an advisor who was helpful and just as hard-working as he expects his students to be. Dr. Michael Roan kept me motivated even when I had doubts about continuing my education. In the end, I am grateful Dr. Roan provided me the opportunity to work under his supervision and contribute to a variety of projects. I also acknowledge Dr. Roan's financial support of my thesis and Master's degree. His constructive criticism these past years is the main reason I finish my education at Virginia Tech with an advanced degree.

I would like to thank Dr. Robert West and Dr. Alexander Leonessa for sharing their knowledge as committee members. I would also like to thank my lab partners Ben Bjerke and Michael Milo for making the office enjoyable as well as a great learning environment. Having others around me who worked hard kept me focused. I specifically thank Michael for this shared knowledge and assistance in the statistical and computer programming fields. I would also like to acknowledge Phillip Zellner, and the MEMS laboratory of the Electrical Engineering Department, whose assistance in providing sensors allowed for the development of a preliminary in-house experimental bearing test bed. I want to thank my father and mother, whose financial support throughout my college career made it possible to experience what I have as an undergraduate along with the motivation to continue my education through graduate school. Their support of Virginia Tech and me made it possible for me to achieve such an education.

CONTENTS

Acknowledgments.....	iii
Contents.....	iv
List of Figures.....	vi
List of Tables.....	viii
Nomenclature.....	ix
1 Introduction	1
1.1 Background and Objectives.....	1
1.2 Outline of Thesis.....	2
2 Introduction to Anomaly Detection and Rolling Element Bearings	3
2.1 What is an Anomaly?	3
2.1.1 Background of Anomaly Detection	3
2.1.2 Condition-Based Monitoring.....	6
2.1.3 Critical Analysis of Anomaly Detection	7
2.2 Background on Rolling Element Bearings.....	8
2.2.1 Structure	8
2.2.2 Fault Kinematics.....	11
2.3 Anomaly Detection in Bearings.....	14
2.3.1 Overview of Industry Practices	14
2.3.2 Time Domain Approaches.....	15
2.3.3 Critical Analysis of Time Domain Approaches	17
2.3.4 Frequency Domain Approaches	18
2.3.5 Critical Analysis of Frequency Domain Approaches	21
2.3.6 Time-Frequency Domain Approaches	21
2.3.7 Critical Analysis of Time-Frequency Domain Approaches.....	23
2.4 Research Gaps and Hypothesis.....	24
2.4.1 Recent Work in Symbolic Dynamics.....	24
2.4.2 Research Hypothesis and Scientific Innovations.....	27
3 Process Background and Vibration Datasets.....	29
3.1 Symbolic Aggregate Approximation (SAX).....	30
3.1.1 Overview of Symbolic Transformation	30
3.1.2 Two-Dimensional Symbolic Transformation.....	33
3.1.3 Critical Analysis of SAX.....	36

3.2	Anomaly Discovery through Markov Models.....	37
3.2.1	Transition Probability Modeling.....	37
3.2.2	Two-Dimensional Transition Modeling.....	39
3.2.3	Critical Analysis of Markov Models.....	42
3.2.4	The Unique Perron-Frobenius Eigenvector.....	44
3.3	Synthetic Bistochastic Data Generation.....	45
3.4	Experimental Fault Data.....	49
3.4.1	Experimental Procedure and Data Acquisition	49
3.4.2	Assumption Validation.....	51
3.5	Numeric Multiple Degree-of-Freedom Model	55
4	Computational Process and Analysis.....	62
4.1	Training and Testing Procedures	62
4.2	Bistochastic Data Performance	63
4.3	Experimental Data Performance	67
4.3.1	Parameter Selection and Analysis.....	68
4.3.2	Distribution Analysis and Regression Modeling.....	71
4.4	Propagating Fault Model Performance	74
5	Critical Analyses and Conclusions	79
5.1	Critical Analysis of Bistochastic Data	79
5.2	Critical Analysis of Experimental Fault Data.....	80
5.3	Critical Analysis of Propagating Defect Data	82
5.4	Research Benchmark and Conclusions	84
5.5	Short-Term and Long-Term Future Work.....	85
	References	87
A	ISO 10816 Vibration Standards.....	95
B	CWRU Experimental Test Conditions.....	96
C	Numeric MDOF SIMULINK Model	97
D	Calculation of Stiffness and Damping Matrix.....	98
E	1D and 2D SAX MATLAB Codes	100
F	2D HMM generation MATLAB Codes	102
G	Experimental Data ANOVA Tables	104
H	Bearing Rig Development	106

LIST OF FIGURES

Figure 2.1. Example of anomalies in a two-dimensional dataset.	4
Figure 2.2. Simulated ECG signal with an anomalous heartbeat.	5
Figure 2.3. (a) Deep-groove rolling element bearing and its (b) cross section.	9
Figure 2.4. Load distribution of rolling element bearing under radial loading.	12
Figure 2.5. Bearing containing all three main localized fault locations.	12
Figure 2.6. Shock pulses in faulty bearing vibration responses.	14
Figure 2.7. Example bearing vibration FFT with a fault frequency at 165 Hz.	19
Figure 2.8. (a) Envelope analysis for bearing fault diagnosis and (b) the associated FFT.	20
Figure 2.9. Example bearing vibration spectrogram in given an outer race fault.	22
Figure 3.1. Proposed bearing anomaly detection architecture.	30
Figure 3.2. (a) Normal bearing signal and (b) Hilbert-based FFT of its envelope signal.	32
Figure 3.3. (a) Partial signal, (b) the distribution of data, and (c) the symbolic representation. ...	33
Figure 3.4. (a) 1D and (b) 2D spectrograms of a normal bearing signal.	34
Figure 3.5. (a) Power spectrum of one time segment in Figure 3.4.	35
Figure 3.6. Symbolic transform of the spectrogram image in Figure 3.4.	35
Figure 3.7. Example Markov Chain model and correlating row-normalized TM normalized.	38
Figure 3.8. Two-dimensional compressed state transformation.	41
Figure 3.9. (a) Normal and (b) anomalous input TM's for bistochastic data generation.	47
Figure 3.10. (a) 1D and (b) 2D synthetically symbolic signals.	48
Figure 3.11. (a) Normal and (b) anomalous data distribution of 2D bistochastic data.	48
Figure 3.12. (a) Normal signal and (b) OR, (c) IR, and (d) RE signals.	50
Figure 3.13. (a) Good and (b) faulty vibration time domain distributions.	52
Figure 3.14. (a) Good and (b) faulty time-frequency data distributions.	53
Figure 3.15. (a) Normal and (b) anomalous symbol distribution in the time domain. (c) Normal and (d) anomalous symbol distribution in the time-frequency domain.	54
Figure 3.16. Rolling element passing over point defect (a) before and (b) after the defect.	56
Figure 3.17. Rolling element bearing MDOF vibration model in the primary radial direction. ..	57
Figure 3.18. (a) CWRU experimental data. (b) Numerical model vibration response.	59
Figure 3.19. Spectrum of the simulated damaged bearing from Figure 3.18 (b).	61
Figure 4.1. (a) 1D and (b) 2D bistochastic data performance.	64
Figure 4.2. Example synthetic signal generated from bistochastic TMs.	65
Figure 4.3. Euclidean NN for (a) normal data and (b) anomalous data.	66
Figure 4.4. (a) 1D and (b) 2D alphabet size analysis for outer race fault.	69
Figure 4.5. (a) 1D and (b) 2D alphabet size analysis for inner race fault.	70
Figure 4.6. (a) 1D and (b) 2D alphabet size analysis for rolling element fault.	70
Figure 4.7. (a) OR, (b) IR, (c) RE fault and (d) normal anomaly distributions.	71
Figure 4.8. (a) OR (b) IR and (c) RE fault regression analysis.	73
Figure 4.9. (a) Fault signal and (b) anomalous score with labeled ISO vibration standards.	75
Figure 4.10. Anomalous scores through time in state-of-the-art methods.	76
Figure 4.11. Frequency domain of propagating anomalous outer race signal.	77
Figure A.1. ISO 10816 vibration severity chart for machine vibration.	95
Figure C.1. (a) SIMULINK model for developing propagating OR vibration signature.	97
Figure G.1. 2D anomalous measure distributions of (a) 0.007 in and (b) 0.021 in faults.	104
Figure G.2. 1D anomalous measure distributions of (a) 0.007 in and (b) 0.021 in faults.	104

Figure H.1. CAD model of current experimental bearing rig.....	106
Figure H.2. Current experimental bearing test bed.....	106

LIST OF TABLES

Table 3.1. Pseudo code for Markov Chain Monte Carlo data generation algorithm	45
Table 3.2. Performance of numerical bearing vibration model with outer race fault.	60
Table 3.3. Performance of numerical bearing vibration model with inner race fault.	60
Table 3.4. Performance of numerical bearing vibration model with rolling element fault.	60
Table 4.1. Ground-truth bistochastic data, using state-based Nearest Neighbors.	67
Table 4.2. Ground-truth bistochastic data, using one-dimensional SAX.	67
Table 4.3. Ground-truth bistochastic data, using two-dimensional SAX.	67
Table 4.4. Ground-truth experimental vibration data, varying spectrogram window sizes.	71
Table 4.5. Anomalous measure given three different confidence intervals.	74
Table 4.6. Condition monitoring comparison results.	78
Table B.1. Normal test parameters and conditions.	96
Table B.2. Faulty test parameters	96
Table G.1. ANOVA table from all data in Figure G.1 (a).	104
Table G.2. ANOVA table from all data in Figure G.1 (b).	104
Table G.3. ANOVA table comparing the three fault locations in Figure G.1 (b).	105
Table G.4. ANOVA table comparing the OR and IR faults in Figure G.1 (b).	105
Table G.5. ANOVA table comparing the OR and RE fault in Figure G.1 (b).	105
Table G.6. ANOVA table comparing the RE and IR fault in Figure G.1 (b).	105

NOMENCLATURE

Variables

α	Contact angle
β	Segment of data in test dataset x
C	Damping matrix
d	Defect size
δ	Length of data segment in x
E	Modulus of elasticity
ε	Load distribution factor
e	Error value
ϵ	Eccentricity ratio
f	Frequency
F_s	Sampling rate
F_T	Total impact force in defect shock
g	Gravitational constant
Γ	Impact coefficient
γ	SAX index values
h	Fluid film thickness
I	Mass moment of inertia
Φ_κ	Markov state neighborhood
J	Total number of elements
K	Stiffness matrix
κ	Two-dimensional neighborhood radius
L	Bearing width
L_{10}	Bearing catalog rating
Δ	STFT window length
λ	STFT lag length
m	Mass
M	Mass matrix
μ	Anomaly threshold value
η	Anomalousness measurement
N	SAX cardinality
n	Natural frequency modes
σ	Markov state representing Φ_κ
Ω	Mass per unit length
$P(f)$	Frequency spectrum
P_r	Pitch radius
$P_N(f,t)$	Symbolized time-frequency power spectrum
$P_N'(f,t)$	State-based time-frequency power spectrum
$P(f,t)$	Time-frequency power spectrum
φ	Diameter
ψ	Angle from maximum bearing load
Π	Transition matrix
π	Individual probability elements of Π

Q	Load on any given bearing element
ρ	Eigenvalues of $\mathbf{I}$
R	Bearing radius
s	Standard deviation
S	Compressed Markov state set
t	Time
τ	Unique, invariant distribution of $\mathbf{I}$
θ	Angle between rolling element and defect axes
V	Velocity
$\forall$	Volume
v_L	Left eigenvector corresponding to $\rho = 1$
w	Angular velocity
w_n	Natural frequency
x	Raw vibration dataset
$\bar{x}$	Computed mean in x
$X(f)$	Fourier series expansion of x
z	Z-score
ζ	SAX window length

Subscripts

o	Minimum
m	Maximum
B	Rolling element
F	Fluid film
IF	Inner fluid film
IR	Inner race
OF	Outer fluid film
OR	Outer race

Abbreviations

$1D$	One-Dimensional
$2D$	Two-Dimensional
AE	Acoustic Emission
$BPFO$	Ball Pass Frequency Outer
$BPFI$	Ball Pass Frequency Inner
BSF	Ball Spin Frequency
CDF	Cumulative Density Function
$CWRU$	Case Western Reserve University
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EDM	Electro-Discharge Machining
EHD	Elastic-HydroDynamic
FE	Finite Element
FFT	Fast Fourier Transform
FTF	Fundamental Train Frequency

<i>FPR</i>	False-Positive Ratio
<i>HMM</i>	Hidden Markov Model
<i>ID</i>	Inner Diameter
<i>MDOF</i>	Multiple Degree-of-Freedom
<i>JTFA</i>	Joint Time-Frequency Analysis
<i>ND</i>	Non-Discriminate
<i>NN</i>	Nearest Neighbors
<i>OD</i>	Outer Diameter
<i>PCA</i>	Principal Component Analysis
<i>PDF</i>	Probability Density Function
<i>PF</i>	Perron-Frobenius
<i>PSD</i>	Power Spectral Density
<i>RMS</i>	Root-Mean-Square
<i>ROC</i>	Receive Operator Characteristic
<i>SAX</i>	Symbolic Aggregate Approximation
<i>SNR</i>	Signal-to-Noise Ratio
<i>STFT</i>	Short Time Fourier Transform
<i>TM</i>	Transition Matrix
<i>TPR</i>	True-Positive Ratio

1 INTRODUCTION

1.1 BACKGROUND AND OBJECTIVES

Anomaly detection is a prevalent area of work in research and in practice. Anomaly detection is especially widespread among mechanical engineering fields, particularly machine condition monitoring. This type of analysis includes the study of machine sensor signals. Increasingly larger signal databases and the constant evolution of dynamic machine systems require the aid of more advanced methods of digital signal processing. As research expands, condition monitoring processes become even faster and with more accurate diagnostics in nondestructive environments. Not only are these processes able to provide real-time analysis of large dataset machine signals, but industries currently implement such technology on systems where accurate condition monitoring is critical. And, with the ever increasing burden of higher production and quality, the availability of extensive research in this field is also necessary.

This thesis attempts to provide further research in the field of condition monitoring by analyzing machine sensor signals from a statistical symbolic perspective. This thesis also implements a generalized approach where a signal processing algorithm is applicable to a range of machine systems, sensor types, and environments. By analyzing from a statistical standpoint, the proposed method challenges current state-of-the-art methods in demonstrating earlier detection of anomalous signals by tracking deviations from ‘normal’ behavior. This ‘normal’ behavior can include installations errors, misalignments, and manufacturing defects as well as associated experimental noise. The objective is to model machine signal response as a statistical pattern recognition problem in the effort to relate defective system conditions with anomalous quasi-stationary behavior. This thesis demonstrates this ability on simulated and experimental rolling element bearing vibration signals; however, the proposed signal processing algorithm is

not limited to this application. Analysis of datasets containing different defect types exemplifies the proposed method's performance.

1.2 OUTLINE OF THESIS

Chapter 2 presents a literature review of anomaly detection. This chapter includes a brief introduction to anomaly detection in general and how these detection methods correlate to industrial condition monitoring. A discussion on rolling element bearings and where anomaly detection and diagnosis methods currently stand in this field follows. An overview of the research gaps in bearing anomaly detection and how symbolic methods fill these voids concludes the chapter. Chapter 3 follows by providing the background and technical knowledge on the specific two-dimensional symbolic method employed in this thesis to detect bearing anomalousness. This chapter includes a background of the intelligent feature extraction method used to discriminate anomalous data from normal data while summarizing the background and critical analysis of the analyzed datasets. Chapter 4 expands on the experimental procedures and the subsequent analysis and results. These results demonstrate the performance of this thesis' proposed symbolic method on bearing vibration anomaly detection. Finally, Chapter 5 includes the critical analyses of these results, the contributions of the proposed method to the condition monitoring field, and concludes with applicable future work.

2 INTRODUCTION TO ANOMALY DETECTION AND ROLLING ELEMENT BEARINGS

2.1 WHAT IS AN ANOMALY?

Anomaly detection is a significant research problem, even as early as the 19th century [1]. An ‘anomaly’ is a deviation from a well-defined normal, or nominal, behavior. Sometimes, these nonconformities are termed outliers, exceptions, or surprises. Anomaly ‘detection’ refers to the discovery of unexpected behavior or the act of finding patterns that do not conform to typical events. Anomaly detection extends to and already has extensive use in areas such as fraud prevention, cyber-security, military surveillance, and machine condition monitoring. Detecting anomalies is important in that outliers are usually statistically significant [2]. Whether there is an unusually large credit card purchase, irregular computer network traffic patterns, abnormalities in medical imaging, or faulty sensor data, detecting unfamiliar activity is critical. An anomalous electrocardiogram (ECG) signal could indicate preliminary evidence of heart failure. Anomalous credit card purchases present the possibility of identity theft. In any case, anomalies can be serious and require timely action.

2.1.1 BACKGROUND OF ANOMALY DETECTION

Anomaly detection techniques are available for a variety of situations and in almost any number of dimensions or domains. Figure 2.1 illustrates a simple example of data containing anomalies in two dimensions. This particular dataset provides two defined normal regions, N_1 and N_2 . Most observations lie in these nominal regions. Regions A_1 and A_2 suggest the point(s) within these fields are anomalous in that they are sufficiently distinct from normal locations. While this is a basic example of visually differentiating between normal and anomalous data, defining these

normal and anomalous groups can be difficult. The exact anomalous/normal boundary is user-defined in many cases and specific to the type of application [3].

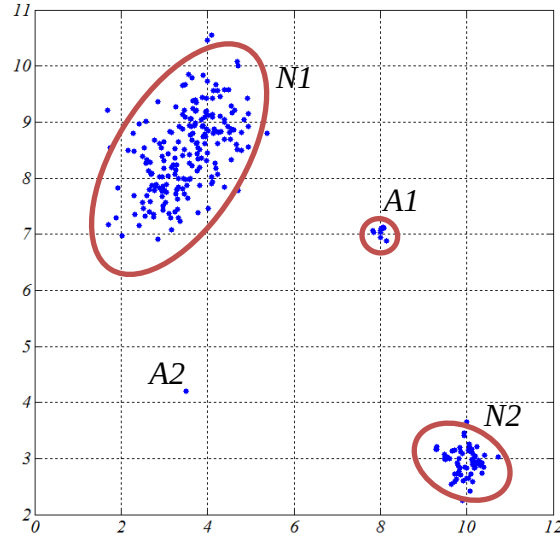


Figure 2.1. Example of anomalies in a two-dimensional dataset.

Generally, anomaly detection is based on machine learning assessment when applied to real-world problems [4]. Machine learning is a program's ability to improve performance over time [5]. This improvement can occur in two types of learning environments. Supervised learning methods are available when there is a known normal dataset. More normal datasets, or training data, improves detection performance. Unsupervised approaches, however, are necessary in applications where there are no labels, or previously known normal regions. In both learning approaches, basic methods require *a priori* assumptions to distinguish anomalies in future observations. Commonly, this is the assumption that anomalies are not central to most data [6]. In general, if an unknown distribution z of data with a dimensionally-dependent density d occurs on the input space Z with respect to a reference distribution μ on Z , anomalies occur where $d(z) > \rho$, where ρ is a fixed threshold level based from μ . A set $\{z < \rho\}$ describes normal observations. Using a simulated ECG signal in Figure 2.2, an anomalous heartbeat clearly occurs at $t = 3$ seconds. 3 normal heartbeats occur from $t = 1$ to $t = 3$ seconds. Many times,

anomaly detection occurs in subsets of data, or windows, where transitions occur naturally [7]. Here, the length of one heartbeat is typical for ECG window length. Parts of the signal densities $d(z)$, where z represents each of the individual heartbeat windows, that extend beyond a user-defined threshold ρ are anomalous, which describe the 4th heartbeat. Window densities lower than the threshold show no evidence of anomalous activity.

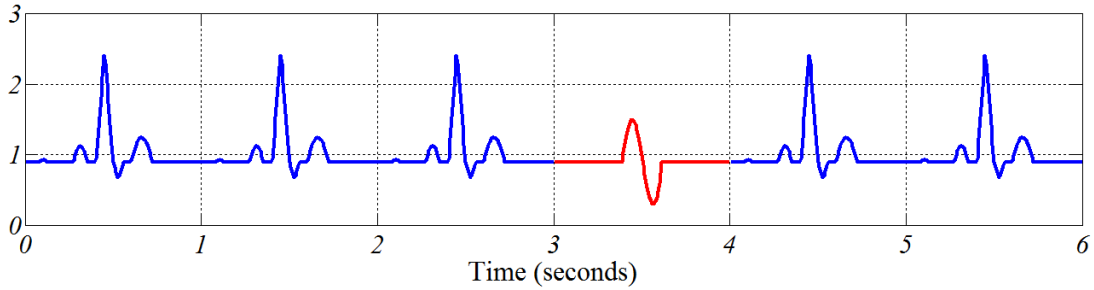


Figure 2.2. Simulated ECG signal with an anomalous heartbeat.

Many anomaly detection practices can detect this irregular heartbeat in the example above, as provided by Karpagachelvi et al [7]. However, the specific area of research, nature and availability of data, constraints, and output requirements influence the technique selection. Broad surveys of general anomaly and outlier detection techniques and their advantages and disadvantages can be found in Chandola et al [2]. Machine learning, or intelligent-based, anomaly detection derives from one of a few generic methods [8]. Classification-based methods, in their most basic form, categorize a test instance from learned models (testing) generated from a set of pre-labeled data observations (training). These can occur in multiple classes, such as Figure 2.1, where two classes are used. Typically, areas of normal and anomalous data are previously known. Bayesian and Neural Network-based analysis [9] are branches of this tactic. Nearest Neighbor-based techniques are common methods where normal data assumes to occur in dense neighborhoods. Anomalous data maintains a certain distance between these neighborhoods [10]. More recently, research in statistical-based techniques is growing, such as

Principal Component Analysis (PCA) [11]. Anomalies are the result of suspicion that certain data is not created by an assumed stochastic model [12]. Normal data occurrences have high probabilities, whereas, anomalies have low probabilities.

2.1.2 CONDITION-BASED MONITORING

Anomaly detection is specifically applicable to mechanical systems. In general, these methods are referred to as condition-based monitoring (CBM) and attempt to detect a dynamic system's deviation from nominal behavior over time. These applications reduce maintenance and prolong machine life. Non-conforming changes in a system's performance can result from self-excitation, as in a gradually developing misalignment in a rotor coupling, or from external stimuli, as in the natural frequency excitation from adjoining machine's vibration. In recent years, large investments occurred in the CBM/structural health monitoring fields, especially in the aerospace and manufacturing industries [13]. For example, General Electric (GE) and Boeing joined to implement CBM on aircraft for embedded health monitoring for a 10-fold increase in real-time performance in 2009. Early detection of anomalous behavior can indicate faulty conditions before time-consuming and expensive repair and downtime.

CBM, or otherwise called predicative maintenance, is a non-destructive and real-time procedure [14]. The benefits of such methods for machine monitoring include [15] [16]:

1. Minimizing ownership cost by lengthening maintenance intervals.
2. Minimizing downtime and safety by reducing component run-to-failures.
3. Increasing component use by maximizing available practical life.

Anomaly detection-based CBM methods, as opposed to specific fault diagnostic techniques, which require human expertise for detection and are described later, have a distinct advantage in early detection of unusual system behavior in that anomalous activity occurs even

before ‘faults’ fully progress. In general, predicative maintenance possesses the following characteristics [17]: (1) Generic algorithm for many machine types; (2) Capability to monitor multiple data types (pressure, temperature, vibration, etc.); (3) Capacity to learn different operating trends over time and detect anomalies from these trends; and (4) Ability to alert incipient defects in a timely manner.

2.1.3 CRITICAL ANALYSIS OF ANOMALY DETECTION

Several factors make anomaly detection in CBM challenging. Identifying normality which identifies all possible nominal behaviors of a system is difficult, as the nature of mechanical components and machinery is dynamic [2]. The most difficult aspect of anomaly detection is determining the boundary between normal and anomalous behavior. Anomalous observations may lie close to the normal boundary, or vice versa. In particular, discerning between anomalous and normal behavior becomes even problematic in components, such as bearings, which may contain one or more of a variety of different ‘anomalous’ conditions (e.g., faults on different component parts at the same time). In more advanced cases, highly anomalous observations start to become more normal in nature (e.g., defects in machinery becoming more physically rounded over time) [18].

As stated before, statistical methods of anomaly detection are becoming increasingly prevalent. Specifically, statistical-based CBM overcomes many of the disadvantages in mechanical systems previously mentioned. Training data is usually less difficult to acquire. Obtaining normality simply implies monitoring data early in the installation of the component/machine or acquiring historical data [19]. Even when manufacturer imperfections or inaccurate installations are present in a system, anomalousness may still develop from this ‘normal’ behavior by tracking system decay over time. This decay includes the possibility of

multiple developing defects. Also, anomaly detection in mechanical systems does not require the prior knowledge of flaws [5] nor the intervention of human experts for detection [20]. Anomaly detection is typically a general approach and, thus eliminates the need for highly accurate and specific system models which may be difficult to attain for all system responses [13]. This thesis incorporates a statistically-based technique, where normality will be known *a priori*.

One common CBM application extends to rolling element bearings, due to the high dynamic load transfers that occur in these machine components as well as high cost of failure and production loss. Rolling element bearings produce distinct anomalous vibration patterns in faulty states, and thus are good candidates for structural health monitoring. The following section provides the necessary background on rolling element bearings and the distinctive vibrational responses produced in anomalous conditions.

2.2 BACKGROUND ON ROLLING ELEMENT BEARINGS

2.2.1 STRUCTURE

Rolling element, or rolling contact, bearings are a generalization of a variety of standardized types: ball bearings, straight roller bearings, tapered roller bearings, as well as many others. Rolling element bearings are further categorized into or, but not limited by, a compromise of deep groove, angular contact, shielded/sealed, self-aligning, or single/double row. In general, rolling element bearings have low starting friction, remain unaffected by operating temperature, and can handle both radial and thrust loads depending on the class [21]. Manufacturer manuals further explain the benefits of the many types. Figure 2.3 below illustrates a basic radial, deep groove ball bearing. This typical bearing structure is an assembly of several parts: An outer ring,

inner ring, set of rollers, and a separator (cage). The most important geometric parameters for vibration analysis are the number of rolling elements, J_B , ball diameter, ϕ_B , pitch radius, P_r , and contact angle, α .

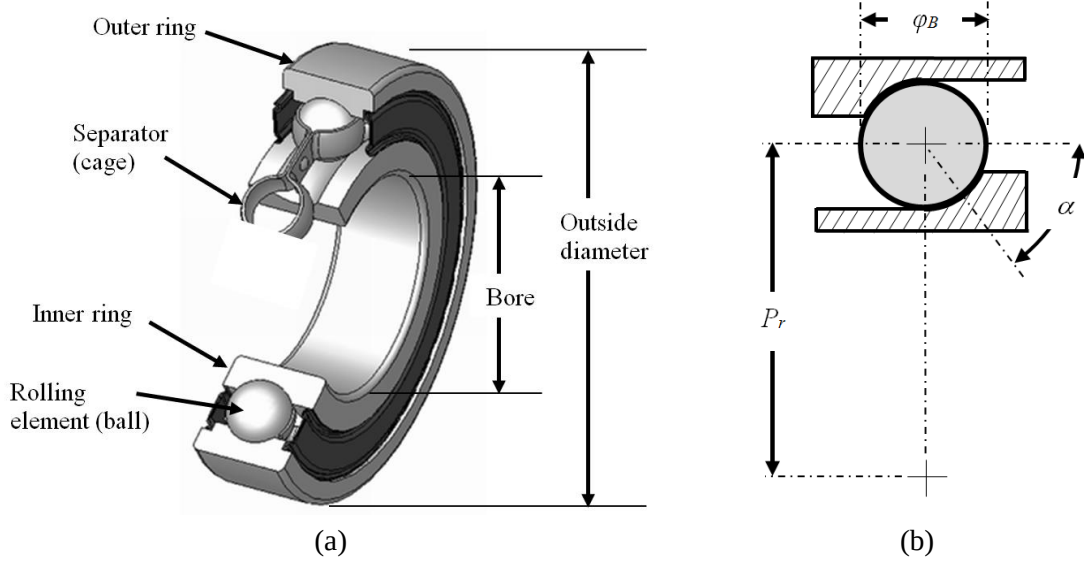


Figure 2.3. (a) Deep-groove rolling element bearing and its (b) cross section.

Metal fatigue is the only cause of failure, assuming the installation and maintenance is performed properly [22]. Common life measures of fatigue failure include the (1) number of revolutions until fatigue onset and (2) the number of hours at a constant angular speed before fatigue onset. Timken, specifically, uses a failure criterion (first sign of fatigue damage) of a spalling size 0.01in^2 , although different definitions exist. Rating life, or L_{10} life, is the common measurement manufactures use to represent bearing life. L_{10} life defines the number of revolutions 90% of bearings will exceed. This 90% threshold represents the stochastic behavior of bearing performance, considering bearing reliability correlates to its probability of survival usually depicted by a Weibull distribution [23]. However, bearings undergo extreme conditions and interrupted maintenance procedures in practice, and other sources of failure do exist. Excessive or reverse loading, overheating, brinelling, contamination, loss of lubrication, corrosion, misalignment, and improper fits are just a few. These all contribute to bearing

defects, are difficult to quantify, and have detrimental and costly effects [24]. In fact, 40% of inductive industrial motor failures result from bearing failures [25].

Bearing failure ultimately produces bearing defects that occur from the flaking of metal particles from the surfaces of the rings [26]. Flaking begins as subsurface imperfections in material weak points which then slowly propagate to the surface. Bearing defects are either localized or distributed classifications. Localized, or point, faults include cracks, pits, and spalls and can occur on the order of nanometers [27]. These point defects typically occur on the races or on the rolling elements where relative motion occurs. Distributed faults include increased surface roughness, waviness, misaligned races, creep, etc. Both fault types increase vibration and noise during operation in the audible frequency range (20 Hz – 20 kHz) [28]. Anomalous vibration is the consequence of rolling elements passing over these defects. By detecting these vibrational impulses over time, an analyst can determine the faulty state of a bearing. For reference, the International Organization for Standardization (ISO) 10816 standard chart for machine vibration is shown in Appendix A. This chart presents the standardized limits for acceptable machine vibration.

Vibration in itself is a three dimensional parameter, with time, frequency, and amplitude. Accurate anomaly detection involves analysis of all three parameters. The typical means of acquiring bearing vibration is through accelerometers mounted near the bearing. Different accelerometers are available, and selection of the type of accelerometer is specific to the application and analysis requirements. Commonly used piezoelectric accelerometers convert mechanical motion to electrical signals, and have good characteristics for shock applications, such as in the case of bearings defects where frequencies can occur from 0 to 20 kHz in

acceleration, 2 Hz to 2 kHz in velocity and 0 to 300 Hz in displacement [29]. Vast research shows bearing failure detection through accelerometer vibration signals is accurate [30] [31].

2.2.2 FAULT KINEMATICS

Despite current advances, rolling element bearings generate vibrations during operation even if no ‘damage’ exists [32]. However, this vibration tends to remain as low amplitude, high frequency, Gaussian-distributed data resulting from operational wear and small manufacturer imperfections [33]. Vibration response also fluctuates from the fact that a discrete number of elements change angular position through time with respect to a reference [34]. Bearing condition monitoring involves the detection of faults or anomalies within this normal, noisy signal. Bearing geometry, fault location, and relative ring speeds determine the specific fault kinematics and anomalous vibration response, assuming no outside sources of vibration.

During operation, load distribution between the rolling elements is not equal [35]. The load on any element, at an angle ψ_B measured from the maximum load Q_m direction (determined by the calculated equivalent load value), is given by:

$$Q_B = Q_{max} \left[1 - \frac{1 - \cos \psi_B}{2\varepsilon} \right]^{1.5}, \quad -\psi_m \leq \psi_B \leq \psi_m \quad (2.1)$$

$$Q_B = 0, \quad \text{elsewhere} \quad (2.2)$$

where ψ_m is the angular limit of the loading, revealed in Figure 2.4 below, and ε is the load distribution factor, determined from the combined axial and radial load components.

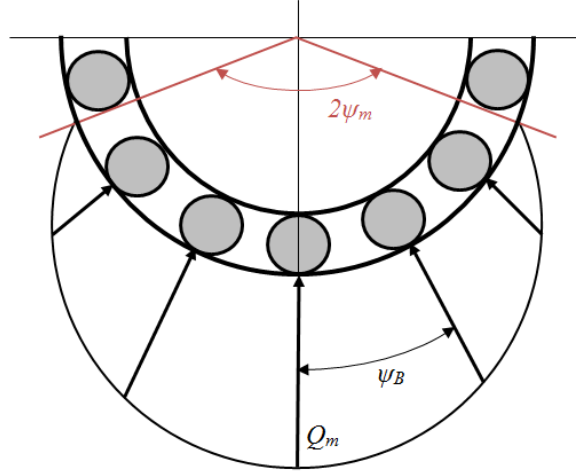


Figure 2.4. Load distribution of rolling element bearing under radial loading.

The location of a fault relative to this load zone will affect the impulse amplitude; with the largest impulse response occurring when faults are in line with the primary radial direction ($\psi_B = 0$). The vibrational response due to faults is also dependent on whether the defect exists on the inner race (IR), outer race (OR), or rolling element (B). Figure 2.5 below illustrates the different positions of localized defects in ball bearings.

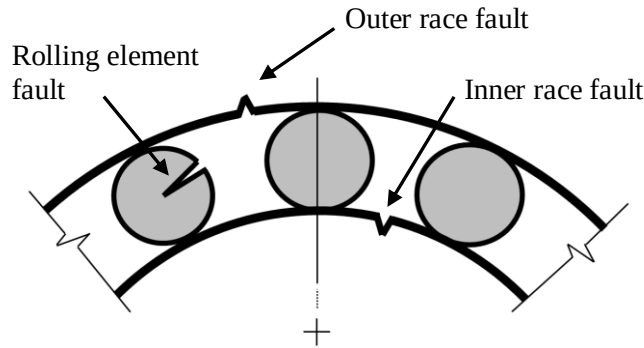


Figure 2.5. Bearing containing all three main localized fault locations.

Assuming both rings are able to rotate, where w_{OR} is the outer ring constant angular velocity in Hz, and w_{IR} is the inner ring's angular velocity in Hz, the different fault frequencies that occur in a rolling element bearing system are found using Equations (2.3) - (2.6). These equations assume all rollers are equal in diameter, in pure rolling contact (no lubrication skid), and no slippage between the bearings and load producing shaft. While these assumptions are

rarely true in a load-carrying bearing showing any wear [21], the equations represent good geometric generalizations of the fault frequencies seen in vibration signals.

$$FTF = \frac{1}{2} \left[w_{IR} \left(1 - \frac{\varphi_B \cos \alpha}{2P_R} \right) + w_{OR} \left(1 + \frac{\varphi_B \cos \alpha}{2P_R} \right) \right] \quad (2.3)$$

$$BPFO = \frac{J_B}{2} (w_{IR} - w_{OR}) \left(1 - \frac{\varphi_B \cos \alpha}{2P_R} \right) \quad (2.4)$$

$$BPFI = \frac{J_B}{2} (w_{IR} - w_{OR}) \left(1 + \frac{\varphi_B \cos \alpha}{2P_R} \right) \quad (2.5)$$

$$BSF = \frac{2P_R}{2\varphi_B} (w_{IR} - w_{OR}) \left(1 - \frac{\varphi_B^2 \cos^2 \alpha}{(2P_R)^2} \right) \quad (2.6)$$

where FTF is the fundamental train frequency in Hz, BPFO is the ball pass frequency on the outer race in Hz, BPFI is the ball pass frequency on the inner race in Hz, and BSF is the ball spin frequency in Hz. If the outer or inner ring is fixed, setting w_{OR} or w_{IR} to 0 Hz respectively is adequate. The specific frequencies in the vibration signal will be made up of one or more of these fault frequencies depending on the fault location. The amplitude at each of the defect impacts is dependent on the phase between these fault-dependent frequencies. Figure 2.6 visualizes this phenomenon for the (a) outer race, (b) inner race, and (c) rolling element fault type with an inner ring rotating at 696 RPM and a fixed outer race. Further discussion on these frequencies in experimental data is shown in Section 3.4.

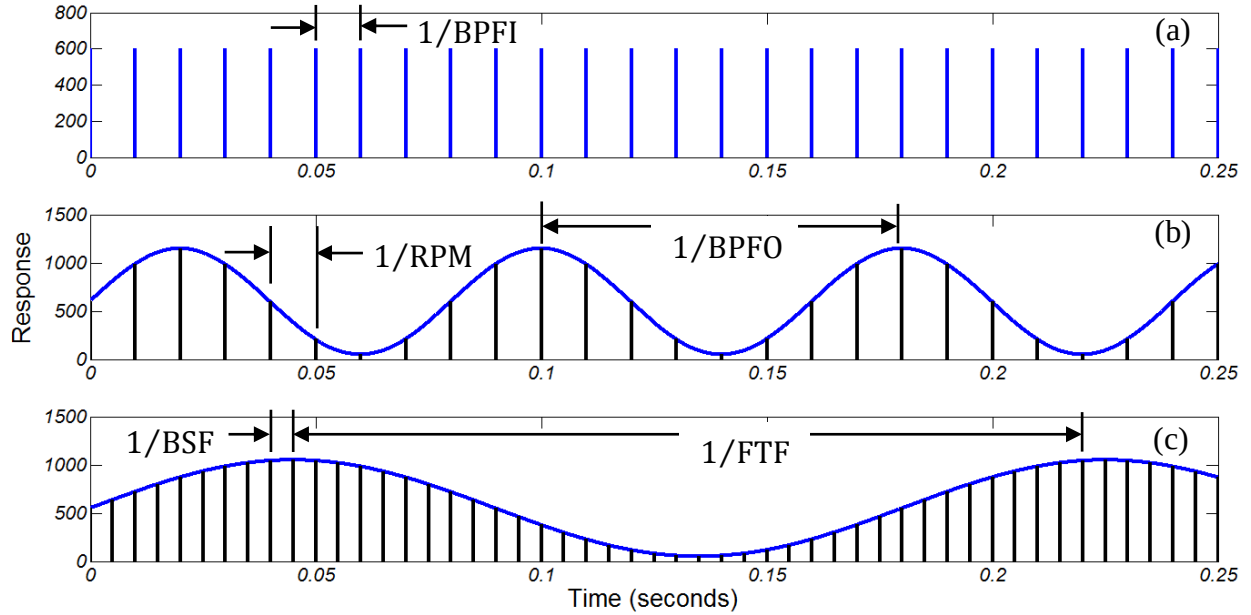


Figure 2.6. Shock pulses in faulty bearing vibration responses.

2.3 ANOMALY DETECTION IN BEARINGS

2.3.1 OVERVIEW OF INDUSTRY PRACTICES

Early detection of faulty bearing conditions prevents catastrophic bearing failure. Especially in high precision/quality environments, reduced downtime and high machine performance is mandatory. Some highly critical situations involve the direct correlation of a bearing's performance to production outcome. Such as in the case of [36], observations reveal that paper mill machine performance directly correlated to bearing operation. Due to the critical nature of bearings, the current use of health monitoring in industry has greatly expanded, with an expected total equipment market of \$2.1 billion by 2015 [37].

Classification of industrial rotating machine vibration monitoring exists in two groups:

- (1) The previously mentioned ISO 10816 preventive maintenance (Appendix A) and
- (2) ISO 13373 predictive maintenance.

The ISO 10816 is simply a guideline by which to gauge the severity of the vibration by its signal amplitudes. This analysis gives no insight into vibration

sources and requires little vibration knowledge. Many times, bearing fault vibration passes as acceptable due to the relatively low amplitude nature of faulty data. Also, maintenance is manufacturer recommended. ISO 13373 are standard procedures for processing vibration data through diagnostic techniques. These processes require a higher level of knowledge in machine vibration to make accurate diagnostics, but permit detection of specific faults and input sources.

An alternative ISO 22096 approach includes Acoustic Emissions (AE) methods. AE techniques monitor frequencies much higher than traditional vibrational techniques [38] by trying to detect the emitted high frequency sound waves generated by transient impact stresses. Clifton and Carino [39] provide that, in some situations, these methods can detect the yielding stage of mechanical failure and thus before weak points become surface faults. Other less common methods include analysis of temperature, electric current, and oil measurements [40]. However, vibration measurements are the most researched and effective for general use of defect diagnosis [41]. While vibration monitoring contains a variety of techniques for different applications, each method is a division of one of three methods: time domain, frequency domain, or a combination of the two.

2.3.2 TIME DOMAIN APPROACHES

Time domain analyses of bearing vibration are simple first-order statistical methods to determine bearing anomalies, and the first known methods of fault diagnosis. Rathbone [42] first presented many original ideas about machine vibration analysis, by including severity criteria based on vibration observations. Time domain analysis could thus connect vibration amplitudes with machine condition and risk of failure. Since then, researchers developed many time-varying scalar measures, based from signal statistics, to quantify anomalous vibration data. The most

used scalar measures include Root-Mean-Square (RMS), standard deviation, kurtosis, crest factor, impulse factor, and shape factor, shown in the following equations.

$$RMS\ value = \sqrt{\frac{1}{J} \sum_{i=1}^J (x_i)^2} \quad (2.7)$$

$$Standard\ deviation = \sqrt{\frac{1}{J} \sum_{i=1}^J (x_i - \bar{x})^2} \quad (2.8)$$

$$Kurtosis\ value = \frac{\frac{1}{J} \sum_{i=1}^J (x_i)^4}{(RMS\ value)^4} \quad (2.9)$$

$$Crest\ value = \frac{Peak\ value}{\left(\frac{1}{J} \sum_{i=1}^J \sqrt{|x_i|}\right)^2} \quad (2.10)$$

$$Impulse\ value = \frac{Peak\ value}{J \sum_{i=1}^J |x_i|} \quad (2.11)$$

$$Shape\ factor = \frac{RMS\ value}{\frac{1}{J} \sum_{i=1}^J |x_i|} \quad (2.12)$$

where $\bar{x}$ is signal mean value of the discrete signal $x(t)$ having J data points. For example, a kurtosis value for an undamaged bearing signal is approximately 3, where vibration data is random in nature. A kurtosis value of greater than 6 is not uncommon for data containing raceway faults, where kurtosis can be considered a measure of data skewness. Particularly, impulse factor and shape factor are effective in Gaussian-based simulated fatigue models; whereas, impulse factor is specifically robust to incipient spalling [30]. Other developments show the use of higher order derivatives as a promising method to mathematically describe the shape of a bearing time signal in faulty stages [43]. In recent years, spike energy time waveforms attempt to isolate fault signatures. Band pass or high pass filters screen the vibration

signal to reject low frequency vibration signals, such as unbalance, misalignment and looseness [44]. Then, the filtered signal passes through a peak-to-peak detector to capture fault frequencies.

2.3.3 CRITICAL ANALYSIS OF TIME DOMAIN APPROACHES

While time domain analysis is a low cost solution to measuring vibration over a wide frequency range, it is less sensitive to emerging defects and provides low diagnosing capability. The low Signal-to-Noise Ratio (SNR) of experimental machine vibration further complicates this phenomenon, as noise buries many of the defective signals. SNR is defined as:

$$SNR = \frac{P_{signal}}{P_{noise}}, \quad SNR_{dB} = 20 \log_{10} \left(\frac{RMS_{signal}}{RMS_{noise}} \right) \quad (2.13)$$

where P is the power in decibels. Current research attempts to reduce noise through Adaptive Noise Cancellation (ANC) [45], by separating the signal from its noise by using a reference noise input to subtract from original vibration source. Still, the high dimensionality of noise contributes to the high variability of system responses, especially when trying to classify signals by peak and range values.

Time domain analysis also demonstrates low adaptability. Outside source vibration easily influence first-order statistical parameters [46]. These first-order statistical methods also present a major flaw in that these scalar quantities typically assume a normal distribution. This assumption is not valid in highly faulty cases [47]. In increasingly larger defects, especially when defect vibration impulses do not expire before the next impulse, this assumption becomes invalid and first-order statistics fails. Kurtosis and Crest factor specifically demonstrate a level of normality in their measures at highly progressed fault vibration. Successive analysis by Heng and Nor [48] demonstrate the ineffectiveness of first-order statistics of faulty data, as progressing fault vibration signatures are not stationary or ergodic (i.e., the probability distribution depends

on time). Frequency analysis attempts to overcome these disadvantages by detecting faulty signals even low SNR conditions, by critiquing signals by known fault frequencies.

2.3.4 FREQUENCY DOMAIN APPROACHES

Frequency domain approaches rely on the fact that localized defects produce distinct characteristic frequencies. Applying a Fourier transform, usually the Fast Fourier Transform (FFT), repetitive vibration signals show as peaks in the frequency spectrum where the repetitive frequency occurs. Higher energy frequencies show as a higher power in the FFT. Given a signal $x(t)$ occurs as a periodic function with period T , the Fourier series expansion $X(f)$ of $x(t)$ is obtained from the Fourier integral:

$$X(f) = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi f t} dt \quad (2.14)$$

where discrete f represents equal spaced frequencies as multiples of the reciprocal of the period T . The power spectrum $P(f)$ is the expected value of complex conjugate of $X(f)$ found by:

$$P(f) = E[X(f)X^*(f)] \quad (2.15)$$

The FFT algorithm is a fast and convenient method for calculating the Discrete Fourier Transform (DFT) of vibration data. The common Welch method of spectral estimation averages the spectrum over the number of windowed data records [49]. Figure 2.7 below illustrates a typical FFT of a bearing vibration signal with an inner race fault frequency of BPFI = 165 Hz. Several fault characteristics exist in this spectrum. Harmonics occur at integer multiples of the fault frequency. The greater number of harmonics, the greater the deterioration [27]. Sidebands also form around these harmonics further indicating increased deterioration. These sidebands typically distance themselves from the harmonic frequencies at $\pm$ the rotating speed, and develop from signal modulation as the defect goes through the load zone. This phenomenon is

especially prevalent in inner race faults where the fault rotates through the load zone. In Figure 2.7, the inner race rotates at $w_{IR} = 65$ Hz. Thus sidebands occur at ± 65 Hz of the original fault frequency and its harmonics. Other considerations include the presence of a raised noise floor which may indicate clearance or looseness issues.

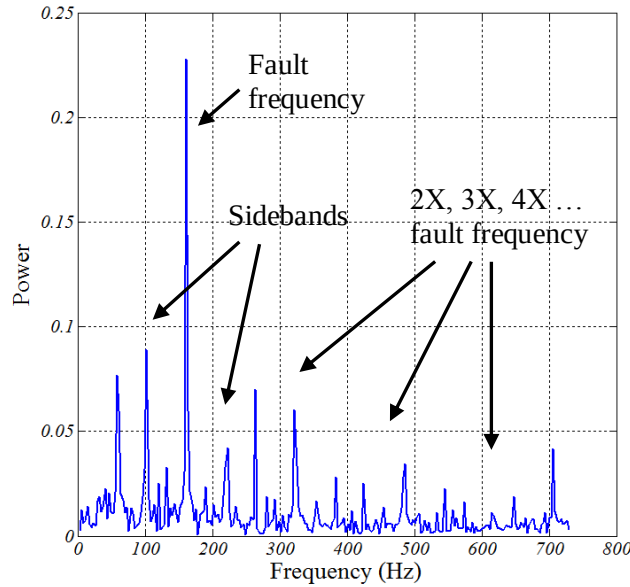


Figure 2.7. Example bearing vibration FFT with a fault frequency at 165 Hz.

An advanced tool in frequency-based analysis is the High Frequency Resonance Technique (HFRT) [31]. Demodulating signals at fault frequencies, involving successive low-pass filters, band-pass filters, rectification, and smoothing, provides an envelope signal with only the fault information. The power of this method comes from the fact that it can detect faulty conditions in low SNR environments or where transducers are not aligned properly [50]. The FFT of this envelope signal discloses information on the defect characteristic frequencies. This HFRT method is also a common tool in industry [27]. Figure 2.8 demonstrates this enveloping process for a bearing fault. Here, the fault frequency again occurs at a BPFI = 165 Hz. Note the envelop signal attempts to model only the high amplitude fault information; however, the FFT of

the envelop signal demonstrates the necessity of a high level of *a priori* vibrational of the fault frequencies.

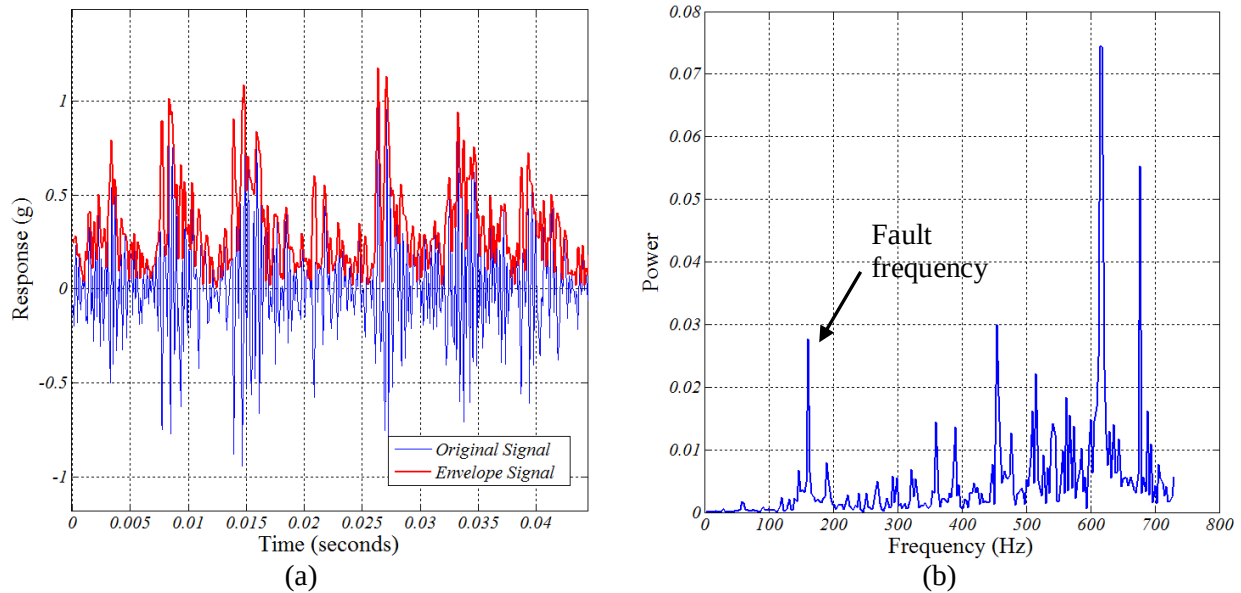


Figure 2.8. (a) Envelope analysis for bearing fault diagnosis and (b) the associated FFT.

Spyridon and Chatzisavvas [51] compare envelope analysis with the raw FFT approach to a vibration signal analysis, and conclude the FFT is capable in finding distinct faults; whereas, envelope analysis provides higher performance in damage onset prediction. Other frequency analysis techniques involve transformations of raw data, such as cepstrum analysis. In the most general terms, the cepstrum domain is the logarithmic result of the FFT and attempts to separate out data echoes. Satyam et al. [52] provides evidence that cepstrum analysis techniques in machinery fault diagnosis using vibration signatures are more accurate and superior to standard frequency analysis. Other transforms, including the Hilbert Transform and Wavelet Transform, help demodulate and refine spectral components of the signal [53]. Mallat [54] demonstrates that use of wavelet coefficients, as opposed to raw data, help raise SNR without the loss of frequency information.

2.3.5 CRITICAL ANALYSIS OF FREQUENCY DOMAIN APPROACHES

Frequency analysis is prevalent and the most common method in vibration monitoring. One of the many advantages includes the availability of highly diagnosable methods. However, as shown in Figure 2.8, direct spectral analysis of these unstable frequencies may not give sufficient information in all cases, thus mandating the use of more advanced algorithms for full analysis, such as the Hilbert-based FFT. Even more advanced detection methods, such as those through harmonic and sideband detection, assumes bearing deterioration is highly progressed. HFRT methods assume the same level of deterioration, and, due to the overlapping of transient components from multiple defects found at the same time, may not identify multiple defects [50].

Further complications exist. Appending structures and sensors will amplify bearing fault impact series [55]. An infinite number of signals can also theoretically carry the same power spectrum [56]. Similarly, frequency analysis requires proper baseline spectrums for comparison. These baselines may be difficult to attain due to the relatively large number of machine operating conditions. The largest drawback of frequency analysis is the loss of time information. One of the ways around the drawback of the lack of time information during FFT analysis is the use of time-frequency analysis, or Joint Time-Frequency Analysis (JTFA).

2.3.6 TIME-FREQUENCY DOMAIN APPROACHES

JTFA is a subset of frequency analysis based on one principle: FFT's are calculated at small enough intervals such that frequency does not dramatically change. The culmination of all segmented FFTs in a signal produces a time-varying power spectrum. Spectrograms are the common display mechanism for these FFT's. Spectrograms are intensity graphs with time, in the x-axis, and frequency, in the y-axis, as variables. To obtain evolution of frequency over time, a

Short Time Fourier Transform (STFT), given by Equation (2.16) allows for FFT analysis using specified window of data.

$$STFT[f, t] = \sum_{-\infty}^{\infty} x[t] * w[\Delta - \lambda] e^{-j2\pi f t} dt \quad (2.16)$$

where $w[\bullet]$ is the time window based on the window location parameters Δ and λ , given by the user-selected data length and time lag values, respectively. The spectrogram, by definition, is the square modulus of the STFT (i.e., $STFT[f, t]^2$). More detailed in the development of the spectrogram is discussed in the following chapter. Figure 2.9 shows an example spectrogram.

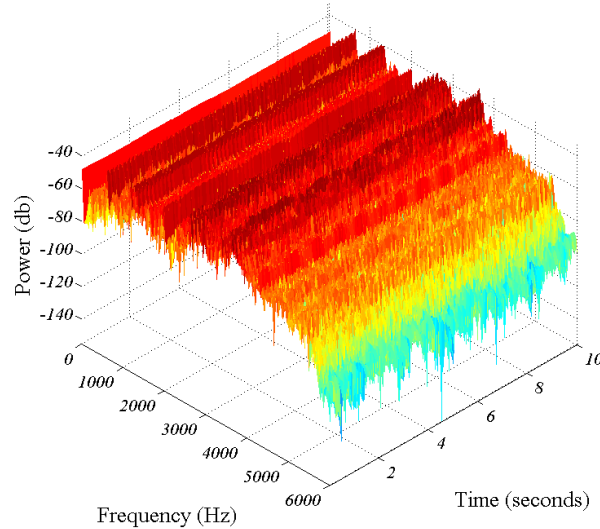


Figure 2.9. Example bearing vibration spectrogram in given an outer race fault.

Time-frequency techniques show potential for detecting bearing problems in more complex rotating machines where the SNR is low and many frequency components are present, as in the common occurrence of multiple defects [30]. Yunlong and Zhenxiang [57] present a time-frequency technique method of decomposing and reconstructing signals based on wavelet analysis, followed by Hilbert Transform-based spectral refining [58]. The experiment results show high effectiveness in revealing the specific details of the fault. Bianchini et al. [59] apply STFTs on various faulty vibration signals in conditions where even time domain scalar quantities

(i.e., Kurtosis) were not reliable, validating the ineffectiveness of using a single characteristic quantity to model the health condition of bearings.

2.3.7 CRITICAL ANALYSIS OF TIME-FREQUENCY DOMAIN APPROACHES

Time-frequency analyses benefits from both time and frequency techniques individually. While time domain analysis provides information about fully developed faults, the performance of such method lacks when first-order statistics of the signal do not change or other machine vibration is prevalent. Frequency analysis alone lacks time information. Anomaly detection is a time dependent environment, and spectrograms provide the advantages that frequency domain information possesses in a time-varying setting. These processes also present the opportunity for real-time analysis even in transient signals, as FFT's are taken in short user-specified time increments [60].

While time and frequency domain analysis can potentially serve as clients for automated, decision-making algorithms for anomaly detection, such as PCA [61], Neural Networks (NN) [62], and even fuzzy logic systems [63], automated detection in the time-frequency domain allows for a greater breadth of research. This extent is due to time-frequency illustrations, such as spectrograms, being represented as two-dimensional images. Feature extraction is simply a case of image processing, where the capability of anomaly detection is broad [64] [65]. Many of these techniques do not require *a priori* assumptions for training data as is needed for traditional approaches nor specific system models [66]. Past research validates the use of decision-making algorithms for anomaly detection in two dimensions. For example, Hu et al. [67] showed PCA-based statistical anomaly detection has an accuracy rate of 98% in detecting anomalies in endoscopic images. Further analysis has been done by Chang and Chiang [68], where anomaly detection in hyperspectral images proved investigation is possible even in large datasets, which is

the case in many mechanical environments. Widodo et al. [69] also confirms the correlation of intelligent machine fault diagnostics of two-dimensional thermal maps to vibrational spectra results. This thesis will expand on current research by analyzing bearing vibration in this two-dimensional vibration spectrum through image processing techniques.

2.4 RESEARCH GAPS AND HYPOTHESIS

This thesis proposes a solution to many of the aforementioned disadvantages of previous methods by processing signals symbolically. Symbolic dynamics is prevalent area of statistical research [70], specifically designed for anomaly detection in experimental signals, where algorithms intelligently learn on signals to detect anomalous and faulty conditions without human response. ‘Symbolic dynamics’ is the general process of data processing and feature extraction for anomaly detection using dimensionally-reduced data. Symbolization, or partitioning, involves transforming raw measurements into ‘symbols’ according to user-selected parameters. Pattern recognition methods mine this symbolically-reduced dataset for information regarding the generating procedure based on an assumed normal stochastic model. This process also allows for on-line analysis and detection of anomalous signals before bearing faults fully progress. Currently, symbolic algorithms are available for many data types and domains, are capable to learn data trends over time, even in transient conditions, and will alert incipient defects in real-time, thus satisfying all condition-based monitoring criteria [17].

2.4.1 RECENT WORK IN SYMBOLIC DYNAMICS

In real-world industrial applications, the critical issue is reliability. Unreasonably missed alarms (i.e., not detecting highly progressed faults) and false alarms (i.e., noise triggered detections) weaken a system’s validity [71]. Unfortunately, experimental observations demonstrate a level

of randomness which can affect this reliability. Much of this variability is a result of low SNR. Symbolic statistics formally incorporate noise. Whether from a sensing device (i.e., measurement noise) or from fluctuations by external inputs (i.e., dynamic stimuli), noise defers from the process at interest and is highly dimensional; whereas, the process at interest is low-dimensional and controlled by low dimensionality trending features [72]. Many times, noise buries this low dimensional response making the process signal almost indiscernible from noise. Symbolic dynamics, however, improves processing and detection of anomalous behavior in such noisy data by incorporating noise characteristics into stochastic models. For example, Cuellar and Binder [73] found adding uncorrelated noise before discretization improves the effectiveness of symbolization.

Symbolic methods expand on slow-time scale anomaly detection. Fast-time scale anomaly detection, or detection where distinct parametric changes occur (i.e., catastrophic failure), is easily understood. The fast-time scale is the local behavior of the dynamic system over which the behavior remains invariant. However, slow-time scale, or detection where subtle changes occur very slowly overtime, is much harder to detect. The slow-time scale refers to the long-term behavior, where dynamic patterns deviate from nominal conditions. Fatigue damage behavior in bearings is essentially invariant on the fast-time scale (i.e., approximately in the order of seconds).

Recent work also indicates that a symbolic dynamic approach is computationally efficient and relatively insensitive to noise, even in chaotic systems where SNR can be as low as 20 dB [46]. Upon discretization of a signal, a string of text represents data trends. Currently, uniform partitioning (UP) [74] and maximum entropy partitioning (MEP) [75] are the most widely used symbolization methods. UP approaches describe signal partitioning methods where equally

spaced bins in the signal's range correspond to different symbols. Averages of data in different data sections correlate to one symbol. MEP provides a partitioning scheme that incorporates variable bin sizes located and spaced to equally distribute signal entropy. Other more advanced partitioning schemes, namely symbolic false nearest neighbors partitioning (SFNNP) [76] and wavelet space partitioning (WSP) [77] optimize symbolic discretization by avoiding data location-dependencies, such as in WSP where the signal's wavelet coefficients are organized into aggregate order. Despite the specific algorithm, the overall goal of any partitioning scheme is to provide an accurate symbolic representation of dynamic systems' quasi-stationary condition, rather than fault specific evolutions [78]. Thus, symbolization has a two-fold effect: procurement of underlying dynamic response, by reducing data variability to only feature trends, and computational speed, by reducing overall dimensionality [79].

A variety of applications specific to mechanical systems employ symbolic analysis for anomaly detection. Analysis and validation of Symbolic Time Series Analysis (STSA), a WSP subgroup, for the detection of anomaly in misaligned flexible rotor system couplings is available for accelerometer data [80]. This particular study reveals that small parametric changes of coupling stiffness in the slow time scale shows accuracy in the predictions of coupling stiffness through anomalous measures. Fatigue damage estimation in polycrystalline alloy structures through symbolic analysis of ultrasonic sensors demonstrate real-time methods for damage monitoring from crack propagation [81]. Chakraborty et al. [82] analyzed aircraft gas-turbine engine vibration anomalies from a symbolic perspective, where vibration responses include multiple input sources and a combination of one or more contributing faults. Even more recently, symbolic filtering progressed to multiple dimensions in itself, as in the partitioning of images [83]. This type of modeling is simply an extension of symbolic time series anomaly

detection to the two-dimensional domain. Image data (i.e., pixels) are the input source for discretization. Subbu et al. [84] experimentally validate one and two-dimensional symbolic dynamics on a fatigue test apparatus equipped with a microscope and ultrasonic flaw detectors. Coupled analysis reveals both one and two-dimensional statistical symbolic models are in agreement and perform well in chaotic systems. Surveys of symbolic dynamic approaches to mechanical system anomaly detection can be found in Chin et al [85].

2.4.2 RESEARCH HYPOTHESIS AND SCIENTIFIC INNOVATIONS

This thesis aims to demonstrate the benefits a symbolic partitioning method provides for bearing anomaly detection in the time-frequency, or two-dimensional (2D), domain. The process here applies a novel method of time-frequency domain symbolic partitioning called Symbolic Aggregate Approximation (SAX). SAX, although a traditional time-domain approach [86], expands into the 2D domain to simultaneously evaluate frequency, time, and amplitude-varying components of vibration allowing for the best scenario of accurate anomaly detection. This analysis is capable of detecting data trends even in low amplitude fault signatures, thus overcoming the limitations of standard first-order statistical analysis.

While this symbolic analysis is robust to noise while reducing data into low dimensionality trends, SAX also provides for an optimal state space environment for Hidden Markov Model (HMM) feature extraction. HMMs mine symbolic signal representations to provide insight into a machine's condition by detecting contextual abnormalities. The 2D HMM approach to anomaly detection not only reduces false alarms, but is able to detect true anomalousness with fewer observations [87]. Since HMM is based on contextual anomaly detection, discovery of anomalousness is not dependent on the impractical assumptions of single input sources, perfect bearing mounting, or steady-state behavior. However, a system's natural

response in transient situations is assumed to be similar to its full quasi-stationary response. Even with this assumption, first-order statistical efforts must allow for transient conditions to phase out.

In general, the process of 2D image analysis for bearing vibration data is relatively recent [88] [89]. The application of machine learning-based feature extraction on bearing vibration images is also a novel endeavor in bearing condition monitoring. However, Hisyam and Aziz [66] specifically assesses STFT on bearing fault diagnostics for on-line condition monitoring and attests to the promises that such algorithms can present. This thesis' proposed process demonstrates an efficient algorithm for finding anomalous trends of data, without the need for human experts in real-time through SAX. This thesis demonstrates the following scientific innovations:

1. Delivery of an intelligent 2D symbolic approach to rolling element bearing anomaly detection through confidence-based HMM analysis.
2. Improvement of detection performance over time and frequency analyses by revealing anomalies in synthetic data where standard first-order statistics fail.
3. Verification of early anomaly detection in condition monitoring applications.
4. Optimization of symbolic parameter selection for bearing vibration data.
5. Provision of a metric for gauging anomaly scores to current standards of acceptable vibration responses.

3 PROCESS BACKGROUND AND VIBRATION DATASETS

This chapter provides the background and technical details of Symbolic Aggregate Approximation (SAX) for anomaly detection along with a discussion on Hidden Markov Models (HMM) as a tool for anomaly discovery. Figure 3.1 below illustrates the general methodology for anomaly detection based on these techniques. A given bearing vibration test signal $x(t)$ is first broken up into individual time segment windows β of equal and specific data point lengths δ . Short Time Fourier Transformations (STFTs), where parameters Δ and λ are the FFT data point window length and overlap values respectively, occur over each of these individual time windows. The result are spectrogram representations $P(f,t)$ totaling the number of time segments in a signal. Conversion of this data in the symbolic domain occurs through SAX partitioning, with N representing the number of symbols used for discretization. A state transformation then constructs a state space representation of this symbolized spectrogram $P_N(f,t)$. The resulting state image $P_N'(f,t)$ correlates to a transition, or probability, matrix (TM), provided through HMM analysis. By comparing the distance between these test TM's and a previous calculated 'normal' TM, the method results in an anomaly score $\eta(\beta)$ at each of test signal's time windows. This thresholding procedure gives an analyst control of the anomaly/normal threshold μ and indirect control of where anomalous occurs in a system.

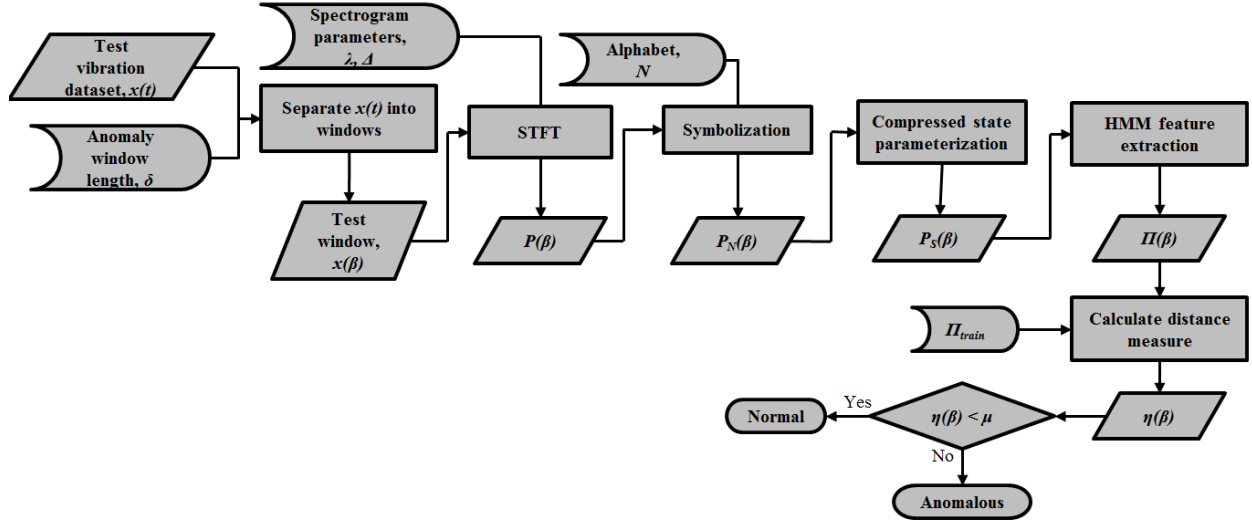


Figure 3.1. Proposed bearing anomaly detection architecture.

3.1 SYMBOLIC AGGREGATE APPROXIMATION (SAX)

3.1.1 OVERVIEW OF SYMBOLIC TRANSFORMATION

Recent published literature regularly uses SAX [90] in fields such as medical and industrial monitoring [91]. Due to the high dimensionality, data reduction methods are ideal in these situations. SAX, in particular, converts data into strings of text by averaging data in window lengths ζ into a discrete number of analyst-specified symbols in alphabet size N . The main attribute of SAX is the assumption that data is normally distributed, where the mean and variance come directly from the raw signal. A bin spans a certain range of data in the signal, where each bin corresponds to a symbol and located relative to the signal mean. Characteristic to SAX, each bin spans a certain range that allows for equiprobable symbol generation. This type of partitioning innately means the range of each bin is not equivalent, where a bin's data range length progressively increases with a more distant relative location from the overall signal mean. For instance, post-partitioning analysis of symbolized data reveals that the total number of each

symbol is equivalent. Because of this type of partitioning, SAX preserves data trends while reducing the overall dimensionality [92]. Due to discretization and indexing of data symbolically, SAX significantly reduces the space requirements to store information as well as computational complexity. Since SAX assumes a normal distribution, the z-score z of any data point x_j is:

$$Z_j = \frac{x_j - \bar{x}}{s_x} \quad (3.1)$$

where $\bar{x}$ is the mean and s_x is the standard deviation of the dataset. Data is then assigned into equiprobable spaced divisions spaced by index values $\gamma_{1,2,\dots,k}$ as in Equation (3.2):

$$\int_{\gamma_m}^{\gamma_{m+1}} f(x) = \int_{\gamma_n}^{\gamma_{n+1}} f(x) = \int_{x=\gamma_m}^{x=\gamma_{m+1}} \frac{1}{\sqrt{2\pi s_x^2}} e^{\frac{1}{2s_x^2}(x-\bar{x})^2} \quad (3.2)$$

This symbolic data conversion is possible in both one and two dimensions. While this thesis presents the application of SAX to the time-frequency domain, an introduction to 1D, or time domain, transformation using SAX is necessary in understanding the specific dynamics of the conversion. This 1D symbolic analysis is also a comparative study by which to gauge 2D performance. Figure 3.2 (a) shows a typical bearing vibration signal acquired through an accelerometer. This type of signal is random in nature. Note that the frequency spectrum of this signal's envelope in Figure 3.2 (b) shows only the frequency of the inner ring speed (29 rev/s).

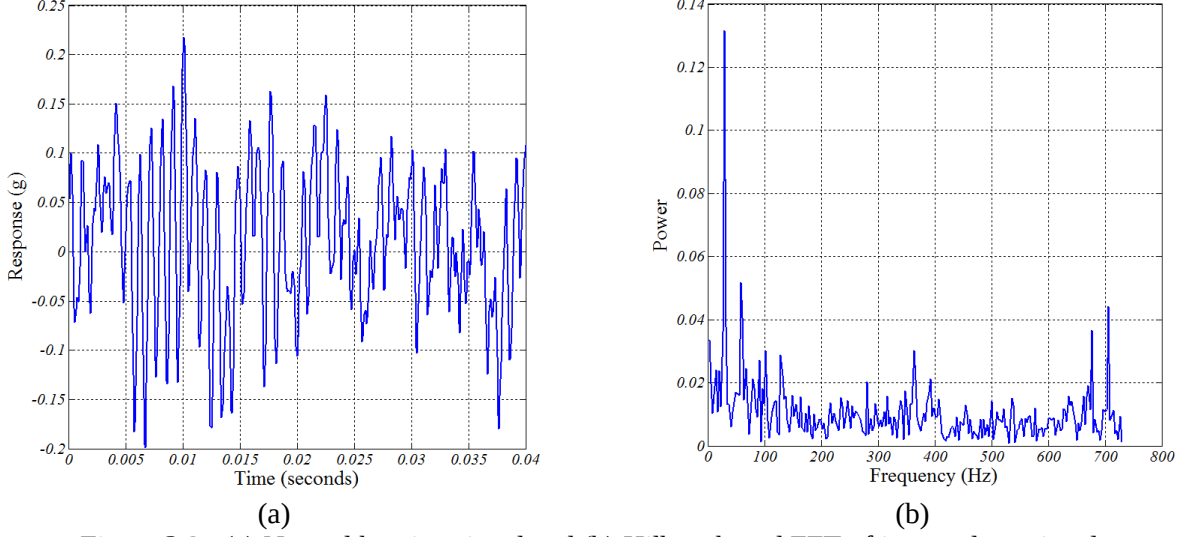


Figure 3.2. (a) Normal bearing signal and (b) Hilbert-based FFT of its envelope signal.

Figure 3.3 illustrates the SAX partitioning scheme on the vibration data from Figure 3.2 (a) given an alphabet size $N = 5$ and window size $\zeta = 0.001$ seconds. The resulting string of text representing the time signal is CCDDCBCB in this brief time segment. Note a dataset mean value of 0 is not a requirement for SAX analysis, due to z-score-based partitioning. While this is strictly for visual purposes, Figure 3.3 presents symbolization based on a selected window size of 1 ms of data, thus each 1 ms of data correlates to one symbol. This string of text characterizes the low-dimensional trend of this section in the first revolution of a normally operating bearing in this condition given the data point window and number of symbols.

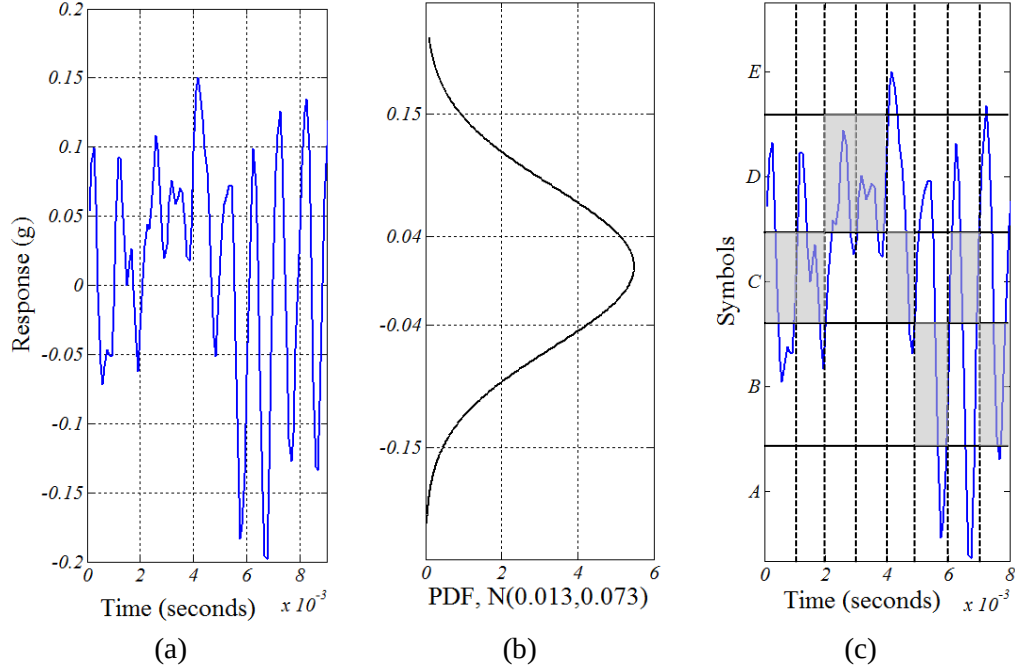


Figure 3.3. (a) Partial signal, (b) the distribution of data, and (c) the symbolic representation.

Continuous analysis of the entire dataset shows this string of text is consistent for this similar beginning portion of each successive revolution of the bearing, as well as a roughly equivalent total number of each symbol *A* through *E* for the entire dataset. While SAX provides good symbolic approximation for good bearing vibration, SAX is not quite as appropriate for faulty vibration, as discussed later.

3.1.2 TWO-DIMENSIONAL SYMBOLIC TRANSFORMATION

This thesis expands on 1D SAX analysis of vibration data by developing a 2D approach to symbol generation. As stated before, analyzing in the time-frequency domain allows for analysis of all three variables of vibration simultaneously: time, frequency, and amplitude. The elements in the Power Spectral Density (PSD) matrix that make up the spectrogram representation of this domain is given as the product of the STFT and a transform $P = |STFT|^2$, where the STFT of the signal comes from Equation (2.16) and where w is the Hamming window to prevent frequency leakage across windows of the STFT. Windowing is necessary in discrete time signals

undergoing discrete Fourier transforms, as even simple waveforms develop non-zero frequencies, or spectral leakages, from incomplete signal frequencies in the discrete time window. Multiple frequencies also interfere with the ability to distinguish between frequencies. The Hamming window demonstrates good characteristics in maintaining a compromise in the ability to distinguish frequencies in highly dynamic noise.

This PSD matrix is the ‘image’ that provides the 2D representation of bearing vibration. Figure 3.4 below illustrates a typical normal bearing vibration signal’s spectrogram with the correlating 2D PSD image.

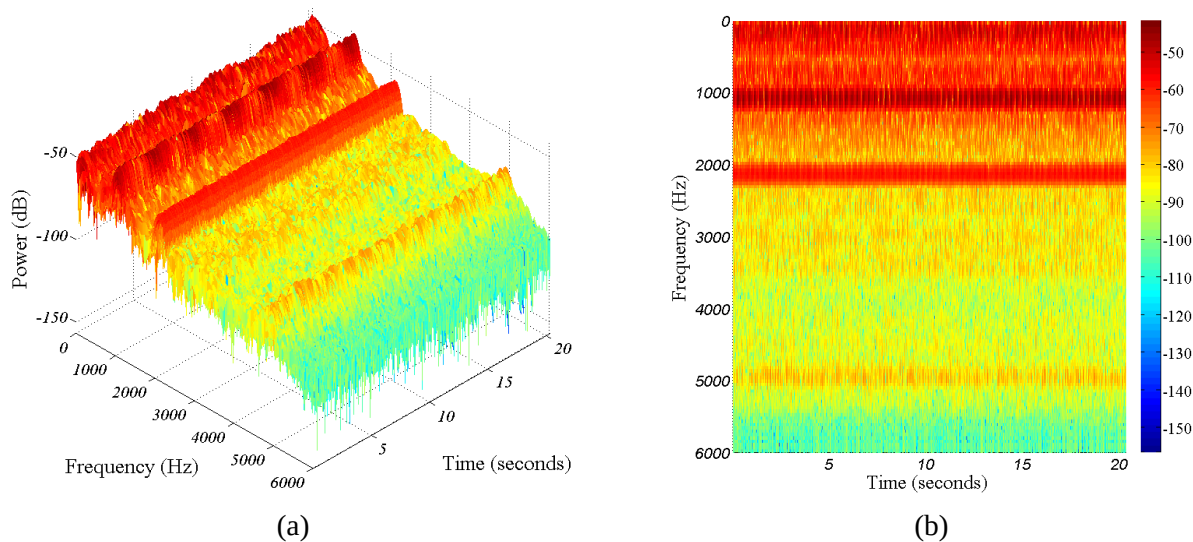


Figure 3.4. (a) 1D and (b) 2D spectrograms of a normal bearing signal.

The spectrogram is the culmination of all STFTs through time. Figure 3.5 below shows the STFT of one time window used in Figure 3.4. Note that this individual time segment power spectrum response roughly correlates to any one of the time segments in Figure 3.4 (a).

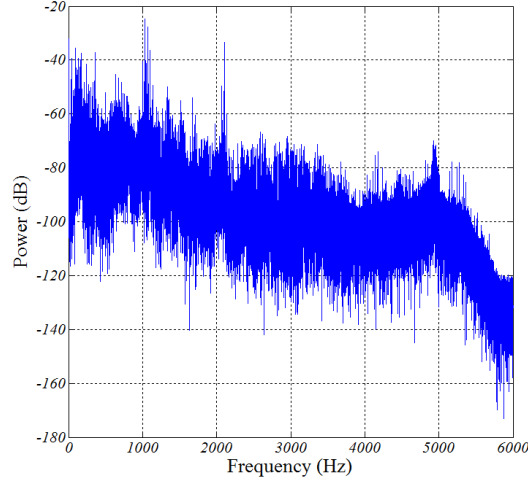


Figure 3.5. (a) Power spectrum of one time segment in Figure 3.4.

2D SAX partitioning involves transforming the power spectral image into symbols in a similar way as the 1D approach. The collective data in the image assumes a Gaussian form, and discretization performs exactly the same. The result is still a 2D image. Figure 3.6 below presents the symbolic domain of good bearing vibration data from Figure 3.4 (a) with a SAX window of $\zeta = 1$ data point and an alphabet size of $N = 8$ symbols. The same SAX parameters used in 1D symbolic transformation are applicable to the 2D domain. If a larger than $\zeta = 1$ data point window applies, windowing involves averaging data in the immediate neighborhood, thus reducing the overall dimensions of the image.

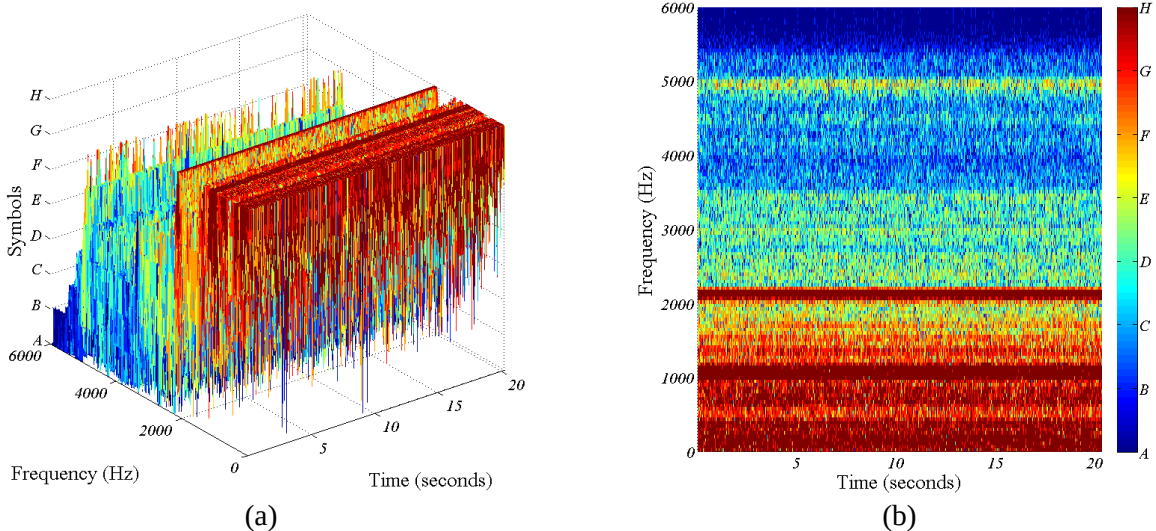


Figure 3.6. Symbolic transform of the spectrogram image in Figure 3.4.

3.1.3 CRITICAL ANALYSIS OF SAX

SAX is especially applicable to mechanical signal partitioning as the dimensionality reduction benefits from the nature of experimental signals. Bearing vibration is relatively Gaussian in nature, and partitioning is done such that trend information is not lost in the process. SAX also offers robustness in both 1D and 2D analysis. Although discussed later, SAX performed on 2D spectrogram images also has the distinct advantage that the distribution does not change between normal and faulty datasets as it does 1D analysis [93]. Lai et al. [94] also proves 2D symbolic dynamics to be effective in chaotic systems over 1D analysis. Other symbolic partitioning methods, particularly uniform partitioning approaches, will discretize in a way that inevitably leads to certain symbols rarely representing any data; whereas, other symbols may correspond to a majority of the data. Statistical analyses in cases when this phenomenon occurs demonstrate a loss of data trends and innately the loss of accurate anomaly detection performance. Recent literature found SAX to be comparable to or better than other dimensionality reduction method, such as the DWT [86]. The main advantages that a symbolic approach such as SAX offers are:

1. Compatibility with low resolution signals (i.e., low SNR), eliminating the need for noise-reduction algorithms and scaling [95].
2. High frequency noise robustness especially in anomaly detection scenarios [95].
3. Real-time processing on inexpensive platforms [81].
4. Improved computational efficiency in high dimensionality data [77] [86].

SAX is also experimentally validated in anomaly detection and machine-learning environments, such as financial data mining [96] and wireless network motif detection [97].

3.2 ANOMALY DISCOVERY THROUGH MARKOV MODELS

Whether in one or two dimensions, this reduced-dimensionality data occupies a discrete set of possible symbols, leading to a state-dependent analysis method for anomaly feature extraction and discovery. Through Hidden Markov Models (HMM), which quantify state dependencies, time-varying anomalous measurements are possible in vibration data.

3.2.1 TRANSITION PROBABILITY MODELING

The symbolic dataset provides a string representing the low-dimensional trends of data. Because of the reduced state-based representation, the environment for statistical analysis of anomalous trends is optimal for Hidden Markov Models (HMM) and Transition Matrix (TM) analysis. Simply, HMM are representations of state-based data that contain a system's state transition attributes given a system comprises of a discrete number of states. 'Hidden' implies the transition attributes are initially unknown and not empirical. These models imply the future responses of a system are probabilities of transitions from one state to another state from the analysis of previously-known data. The term 'Transition Matrix' is the mathematical form of the HMM, or Markov Chain. These transitions between system states form a discrete set of transition probabilities of order N^2 , compiled as a square transition matrix. An example transition probability between two states A and B is computed using Bayes' Theorem, shown in Equation (3.3).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.3)$$

where $P(\bullet|\bullet)$ is the probability of one state given another. Markov Chain generation involves observing data, or sufficient *a priori* knowledge of the data's conditional dependencies. Figure 3.7 (a) below displays a typical Markov Chain model with $N = 3$ states. Figure 3.7 (b) portrays

the TM of a 3 state system. Note that the TM is a 3 x 3 matrix where each cell in the matrix represents the transition probability from one state to the next.

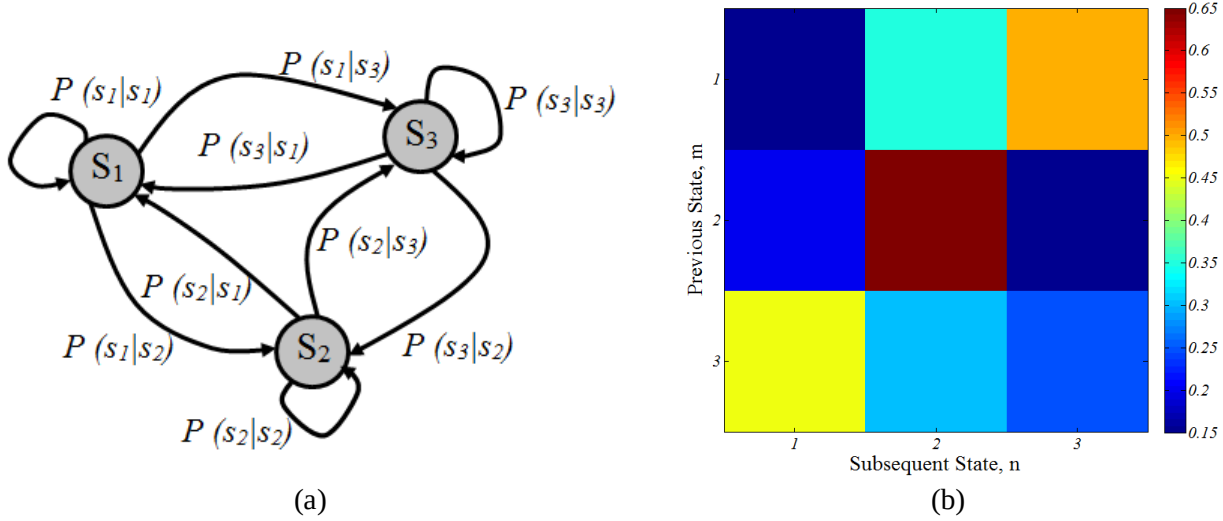


Figure 3.7. Example Markov Chain model and correlating row-normalized TM normalized.

A Markov Chain problem has the following properties [98] [99]:

1. All observations must be in exactly one defined and discrete state at all times
2. Objects will move from one state to the next according to the transition probabilities, which only depend on the current state
3. Transition probabilities do not change over time

Property three is not a general Markov Chain property, only for homogeneous cases. Anomaly detection in machine vibration is a homogenous case and fulfills this assumption. With a given state set $S = \{s_1, s_2, \dots, s_N\}$ made up of all possible states in an alphabet size N , a Markov model is a $|S|^2$ size stochastic matrix construction given by the TM $\Pi = [\pi_{ab}]$ where $\pi_{ab} = P(s_b|s_a)$ is the probability of state s_b being transitioned to given the current state of s_a . Π is given by:

$$\Pi = \begin{bmatrix} P(s_1|s_1) & \dots & P(s_N|s_1) \\ \vdots & \ddots & \vdots \\ P(s_1|s_N) & \dots & P(s_N|s_N) \end{bmatrix} \quad (3.4)$$

TMs also exhibit basic properties. Each element in a TM is a probability. Therefore, each number must be between 0 and 1, inclusive. The elements of each row of the matrix sum to 1, because of the Markov Chain property 2 above. Finally, the TM must be square, since each state corresponds to a row and a column. Generation of this TM is fairly straightforward, accomplished by tallying the total number of each unique state transition in the data and then normalizing the summation matrix. This normalization coefficient is derived by:

$$T_{ab} = \frac{\hat{T}_{ab}}{\sum_N \hat{T}_{ab}} \quad (3.5)$$

such that any valid value of a , $\sum_N T_{ab} \equiv 1$. The Law of Total Probabilities, given in Equation (3.6), relates marginal probabilities to conditional probabilities by stating that the probability of being in state A is dependent on the probability of state B :

$$P(A) = \sum_N P(A|B)P(B) \quad (3.6)$$

With sufficient data, it is possible to generate a complete stochastic TM, such that all transition probabilities are nonzero (i.e., all transitions have been observed to occur at least once in the dataset) and is termed a fully populated TM. This TM is the underlying model for anomaly detection, where anomalousness is the distance between a normal TM and the test TM. In bearing vibration analysis, an anomalous measurement is the distance between the previously known normal TM for good vibration data and a test signal's subset TMs. While this is simple for 1D approach (i.e., successive symbol transition are modeled), TM generation and probability modeling in two-dimensions is slightly more complex.

3.2.2 TWO-DIMENSIONAL TRANSITION MODELING

1D TM models make the assumption that the symbolic time domain signal symbols are, in fact, Markov states. While symbolic signals are commonly analyzed through HMM and easily

conform to Markov state assumptions, HMM analysis in two dimensions must incorporate more extensive approximations. Local neighborhoods centered at symbolic pixel (i,j) in a symbolized spectrogram are the Markov states, as opposed to the pixel element itself. The definition of a neighborhood at pixel location (i,j) is:

$$\Phi_\kappa = \{(i',j') \in X_N(f): \max(|i - i'|, |j - j'|) \leq \kappa\} \quad (3.7)$$

where κ is the neighborhood radius and the PSD symbol at (i,j) is the finite symbol specific to the symbolized spectrogram image $P_N(f,t)$. A Markov state $\sigma_{ij} = P_N\{\Phi_{K-1}(i,j)\}$ defines all possible neighborhoods at a pixel location (i,j) , where the finite state set $|S| \leq |N|^{| \Phi_{K-1} |}$ bounds the total number of states in the spectrogram symbolic image. This inequality corresponds to the absence of a specific state. Grimmett and Stirzaker [100] further describe the Markov inclusion principles for state representation of a neighborhood of symbols or other finite states. Namely, Markov states in two dimensions must adhere to two distinct assumptions:

$$P(P_N\{i,j\}) > 0 \quad (3.8)$$

$$P(\sigma_{ij} | P_N\{i,j\}) = P(\sigma_{ij} | P_N(\Phi_K\{i,j\})) \quad (3.9)$$

This ensures probability of a symbol depends only on the arrangement of symbols in the immediate neighborhood, providing for analysis as a Markov Random Field (MRF). However, this Markov neighborhood can become complex and highly dimensional. Thus, a compressed state set S_c is generated to mask multiple and distinct Markov states σ as equivalent states. Neighborhoods $\Phi_{K-1}(i,j)$ and $\Phi_{K-1}(i',j')$ are equivalent states if the occurrence of each symbol in the neighborhood is equal. The occurrences of symbols, not the spatial arrangement, thus define the Markov state. Where $|N| = 3$ and $\kappa = 2$, the state merging method reduces the number of states from $|S| = 3^9$ to $|S_c| = (|\Phi_{K-1}| + |N| - 1)! / |\Phi_{K-1}|! + (|N| - 1)! = 55$. Two states s_a and s_b are equal if:

$$R_{s_a}(\{N\}) = R_{s_b}(\{N\}) \quad (3.10)$$

where $R \cdot (\{N\})$ is the number of occurrences of any given symbol in alphabet N in a neighbourhood $\Phi_{\kappa-1}(i,j)$. Figure 3.8 illustrates this 2D state transformation. This example displays a state merging transformation where a 3 x 3 size pixel neighbourhood ($\kappa = 1$) is a state, where a state is the total number of symbols in this neighborhood and not the specific arrangement. Here, the upper right 3 x 3 quadrant represents a state S_1 defined by 5 a 's, 2 b 's and 2 c 's. Other symbol arrangements in a separate neighbourhood are an equivalent state to S_1 if the number of each symbol in that neighbourhood matches these totals.

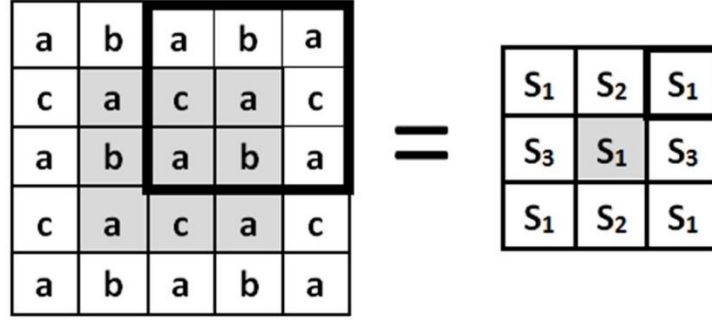


Figure 3.8. Two-dimensional compressed state transformation.

From here, TM generation is relatively the same. However, instead of a state being dependent on a previous state as in one dimension, the TM in two dimensions is developed by neighborhood state pairs. Using the same TM in Equation (3.4), the probability between each state is now in an altered form of Bayes' Theorem where $\pi_{ij} = P(s_a|s_b)$ is the probability of state s_a being a *neighbor* of the state s_b , where the arguments s_a and s_b are states within the compressed state set S_c . Modeling transitions between neighboring states results in a loss of the spatial arrangement of the PSD and instead emphasizes the 2D trends. From the example in Figure 3.8 above, the central highlighted state S_1 in the transformation image includes transitions to S_1 , S_2 and S_3 . 4 (S_1-S_1), 2 (S_1-S_2), and 2 (S_1-S_3) transitions occur. Thus, the TM will be a square matrix

of size $|S_c| \times |S_c|$. For $|N| = 3$ and $\kappa = 1$, the Markov TM will be a 55×55 sized matrix to define all possible state transitions. Adequate *a priori* knowledge of the image data's conditional dependencies is still necessary for Markov model generation. Even in two dimensions, TM's are row normalized (i.e., $P(s_a|s_b) \neq P(s_b|s_a)$). For this paper's analysis, a neighborhood where $\kappa = 1$ remains constant. Therefore, at every symbolic point (i,j) , the κ -neighborhood corresponds to a particular Markov state $\sigma = \{S_c\}$. Other neighborhood values showed negligible performance gains at the cost of significantly higher computation.

3.2.3 CRITICAL ANALYSIS OF MARKOV MODELS

Stochastic modeling is based on likelihoods or probabilities of occurrences; whereas, deterministic approaches predict single outcomes. Machine failure, especially in bearing performance, is a stochastic outcome where failure is the random variable. Markov models look at events and analyze tendency on one event followed by another. In the case of machine failure, this modeling equates the probability of a machine failing given a current system state. Incorporated with SAX partitioning, this analysis can be done even in low SNR environments.

HMM is also statically engineered for anomaly detection, even in systems where first-order statistics fail. In the example of a coin flip, the probability of flipping heads is 50% despite whether the current flip was heads or tails. Assuming 100 coin flips with 50 heads and 50 tails, first-order statistics, such as histogram analysis, initially classifies this coin as 'normal'. However, anomalousness may still be present. If the coin flip occurrences were *Heads-Tails-Head* through all 100 flips, the coin seems as if the probability of heads is much higher if the current flip is tails, and vice versa. Being able to detect internal transition probabilities, not just overall probabilities, gives the analysts the opportunity to detect anomalous occurrences within

data that first-order statistics cannot [101]. In the case of the symbolic vibration signal, this is the probability of the next state to occur given the current state.

Markov methods are powerful in predictive maintenance. Bearing vibration in particular is a system exhibiting state dependencies [80]. The significant advantages over other reliability modeling techniques include:

1. Given the complex dynamics of many mechanical applications, HMM classifies system response as one simple $N \times N$ matrix.
2. Anomalouslyness is a measure of differing probabilities, not deterministic outcomes, thus degradation is on a continuous scale, or through ‘partial’ failure.
3. The only real assumption is that the steady state conditional probability of the next state is only dependent on the previous state [102].

Because of advantage 3 above, the SAX assumption of normally distributed data is not a full requirement for all data [103], only that state transitions only be dependent of the previous state, or symbol. This is important in the analysis of power spectrums.

The major drawback in HMM is the number of states. The number of states increases as the size of the system increases. However, the number of states in vibration analysis is user-controlled in the symbolic transformation stage. While some attempts have been made to mathematically optimize the number of states [104], this is ultimately dependent of the signal type, SNR, and output requirements. And, with a κ -neighborhood length of 1, each pixel element is its own state, further reducing the number of possible states in the SAX-conversion. While HMM has only fairly recently been applied to two dimensions for anomaly detection [105] [106], this thesis also attempts to bridge this research gap in machinery monitoring by applying HMM to symbolic data for anomaly detection in PSD images.

3.2.4 THE UNIQUE PERRON-FROBENIUS EIGENVECTOR

Given that the TM Π of a bearing vibration signal is nonnegative and ergodic ($\Pi \in R^{N \times N}$, $\Pi^k > 0$), or for some k there is a path of length k from every node to every other node, there is an left eigenvalue ρ of Π that is real and positive with left and right eigenvectors. The ρ has a multiplicity of 1 and is called the Perron-Frobenius (PF) eigenvalue of Π [107]. The associated left eigenvector is termed the PF eigenvector, is unique, and found by Equation (3.11):

$$v_L \Pi = \rho v_L \quad (3.11)$$

where v_L is the left eigenvector satisfying the above equation. TM's are stochastic and follow these assumptions. Since the TMs are also regular, there is a unique and invariant distribution τ which satisfies $\tau > 0$. The eigenvalue 1 is simple and dominant, so the TM can simplify into τ no matter what the initial probability distributions. In other words, the distribution of a regular Markov Chain always converges to this unique invariant distribution τ . The 2D $|S_c| \times |S_c|$ TM thus converges into the left eigenvector $v_L = \tau$ of the TM with the dominant PF eigenvalue $\rho = 1$.

The anomaly measure between two Markov TMs, or the difference between the test and trained TM in the case of detecting anomalousness in mechanical signals, is the simple l_1 norm distance between these unique invariant distributions τ_{test} and $\tau_{trained}$, shown in Equation (3.12).

$$d(\tau_{test}, \tau_{trained}) = \|\tau_{test} - \tau_{trained}\| = \sum_{i=1}^N |\tau_{test}(i) - \tau_{trained}(i)| \quad (3.12)$$

where $\tau_{test} = \{\tau_1, \tau_2, \dots, \tau_n\}$ and $\tau_{trained} = \{\tau_1, \tau_2, \dots, \tau_n\}$ are the vectors corresponding to the two left eigenvectors with n -dimensional real vector space. The $d(\bullet)$ is the distance and the resulting scalar measure represents the total anomalousness nature of the signal [108]. Since the elements of the eigenvector are between 0 and 1, the anomaly score is normalized by the total number of

elements n in the vector, as to prevent anomaly scores outside 0 and 1, inclusive; zero being completely normal and one as completely anomalous.

3.3 SYNTHETIC BISTOCHASTIC DATA GENERATION

This section along with the remaining section in this chapter outline the processes and background of the datasets used in determine the performance of 2D SAX and HMM anomaly detection. This section specifically details the development of the synthetic symbolic data. Often, state transition probabilities can generate data for Bayesian analysis through methods such as Markov Chain Monte Carlo (MCMC). In such ways, synthetically-generated data mimics the input stochastic transition data, so that the conditional transition probabilities between states are nearly identical to different dataset's TM. By observing a data's TM in a weighted random walk, sequential data points are derived from the sampling of state-dependent histograms. Pseudo-code for such a sampling process is given in Table 3.1.

Table 3.1. Pseudo code for Markov Chain Monte Carlo data generation algorithm

$TM = (\text{Transition Matrix}),$	$S = (\text{Number of States})$
$m = (\text{System State to Start})$	$J = (\text{Length of Generated Data})$
for $j = 1$ to J	
$breakPoints(1,1) = 0$	
$breakPoints(2,1) = TM(j,1)$	
for $s = 2$ to S	
$breakPoints(1,s) = breakPoints(2,s-1)$	
$breakPoints(2,s) = breakPoints(1,s) + TM(j,s)$	
$randomDraw = RAND \in (0,1)$	
for $p = 1$ to J	
if $breakPoints(1,p) < randomDraw \leq breakPoints(2,p),$	
break	
$nextState = p$	
$m = p$	

The Cumulative Density Function (CDF) for a prior state m is the basis for calculation of the matrix $breakPoints$, and $nextState$ is returned to the calling function. This $nextState$ becomes

the prior state for the subsequent transition, and so on, until the dataset of size J is fully generated, enabling ground-truth analysis of detection performance. This method also expands into two dimensions. Since the analyst-selected neighborhood parameter is $\kappa = 1$, the probability of the neighborhood state, instead of the next 1D state, is dependent on the input TM.

Traditional analysis methods such as frequency analysis or histogram analysis are not well-suited to extract information from high-noise data. An even less advantageous case would be if the data resembles white noise. This random state-dependent data fits the criterion:

$$P(a) = P(b) = P(c) \dots = P(N) \quad (3.13)$$

This equiprobable quality is observed in nonrandom deterministic data, in cases where the total probability of each state is equal, but not the probability of state transitions. A bistochastic (i.e., doubly stochastic) matrix is one such case, and will be used to demonstrate that the methods developed in this thesis are robust for low SNR data. Algorithm 1 of [109] details an iterative method for forming nonrandom bistochastic matrices from any nonnegative, real TM. This process normalizes each row and column to unity in an iterative manner. Given the input $N \times N$ matrix Π , this method will:

1. Normalize each row of Π by dividing it by the sum of its elements.
2. Normalize each column of Π by dividing it by the sum of its elements.
3. Stop execution if the matrix Π is normalized to a specified precision.

The Summation of Absolute Differences (SAD) for each row and column is the metric for user-specified precision of the matrix, shown in Equation (3.4):

$$\epsilon = \sum_{a=1}^N ARG_ABS\{\Pi_{a,b} - 1\} + \sum_{b=1}^N ARG_ABS\{\Pi_{a,b} - 1\} \quad (3.14)$$

where error term, ϵ , is the desired degree of accuracy. For reliability, this threshold is $\epsilon < 0.01 * N^2$, so that the mean bistochastic error is less than one percent. Additionally, when generating this matrix, it is desirable to choose nonrandom structures for the state transition probabilities. These structures are the separate and unique bistochastic TM's for both normal and anomalous data. The structure of normal data is governed by $\Pi_{a,b} = 0.75 + RAND \in [-0.2, 0]$ and anomalous data is governed by $\Pi_{a,b} = 0.1 + RAND \in [-0.5, 0.05]$. This schema provides for a state structure where the probability of a state transition back to own state in normal behavior is roughly between 55% and 75% and approximately 10% from any state to any other in anomalous conditions.

Figure 3.9 below illustrates both the normal and anomalous TM inputs into the generating algorithm. Both the row and column independently sum to 1 with $< 1\%$ error rates ($\epsilon = 5.8575(10)^{-5}$ and $4.8766(10)^{-5}$, respectively). Here, the alphabet size is $N = 10$.

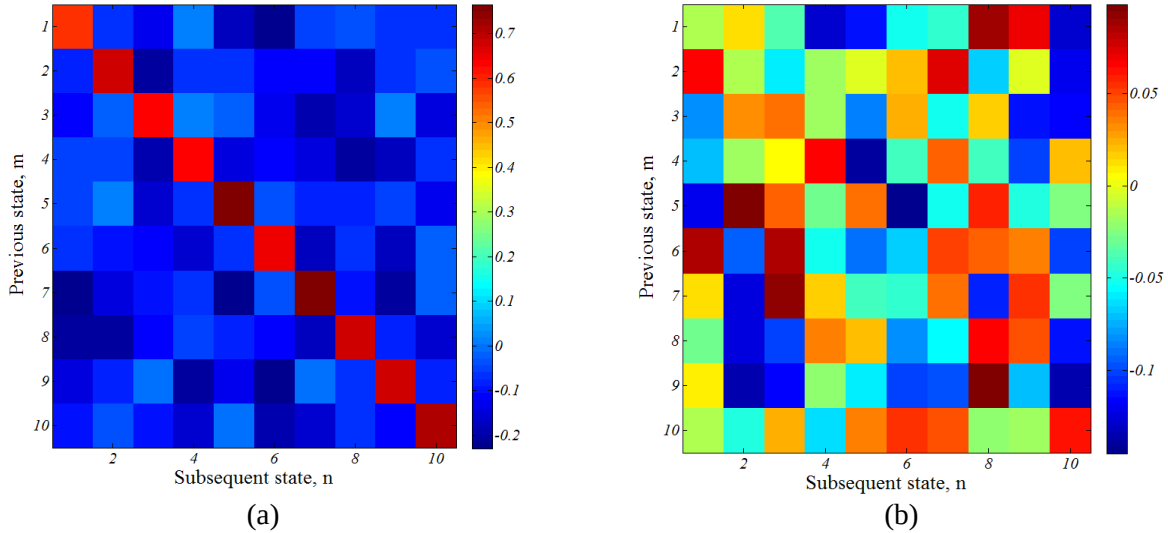


Figure 3.9. (a) Normal and (b) anomalous input TM's for bistochastic data generation.

A sample 1D state signal and 2D state image where the first 500 points of data are generated from the normal TM and the second 500 data points generated from the anomalous TM are shown in Figure 3.10. For reference, the mean and standard deviation of normal 1D data

is 5.496 and 2.873. For anomalous data, this is 5.502 and 2.872. In the 2D domain, these were 5.494 and 2.870 for normal data and 5.502 and 2.873 in anomalous conditions. Figure 3.11 below illustrates the histogram analysis of the above normal and anomalous data in two dimensions using $1(10)^6$ transitions and 10 bins.

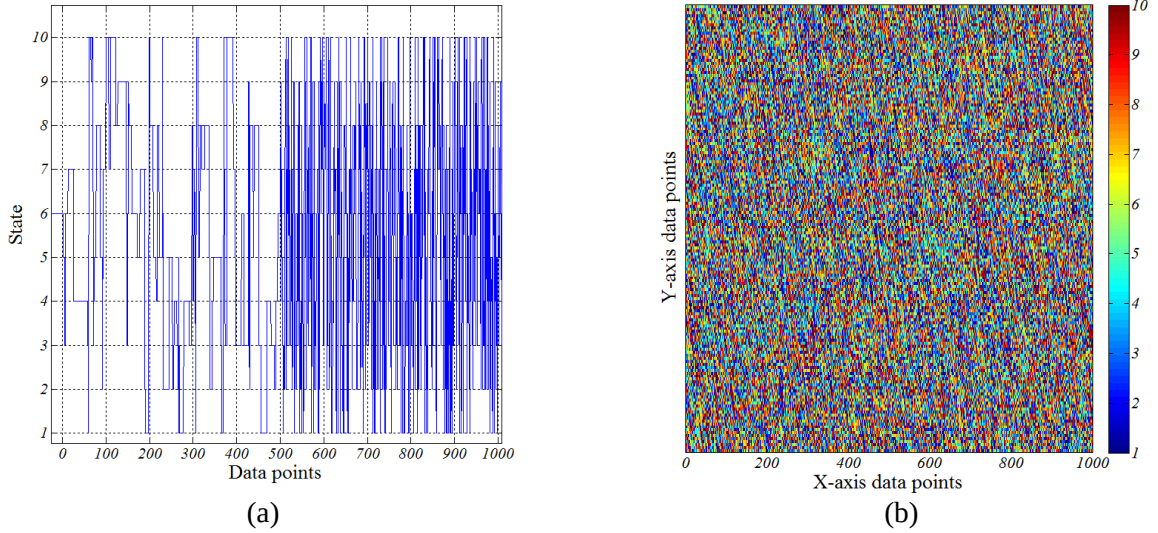


Figure 3.10. (a) 1D and (b) 2D synthetically symbolic signals.

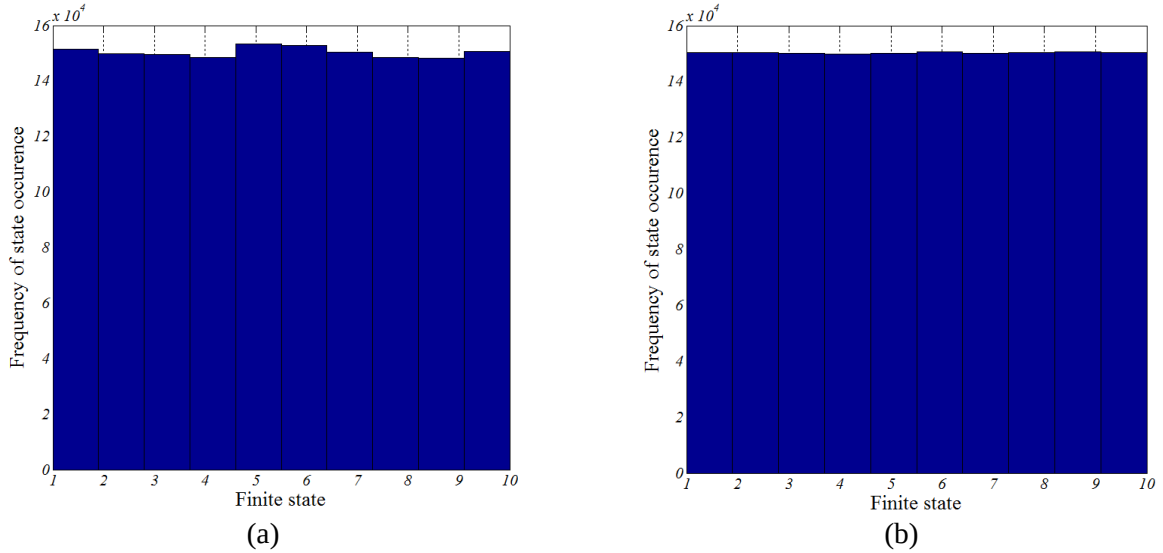


Figure 3.11. (a) Normal and (b) anomalous data distribution of 2D bistochastic data.

Over these large datasets, both normal and anomalous data occurrences will be equal and are functionally indistinguishable using first-order distribution analysis. As the same input TMs generated both 1D and 2D synthetic datasets, the distributions will be concurrent in both as well.

3.4 EXPERIMENTAL FAULT DATA

3.4.1 EXPERIMENTAL PROCEDURE AND DATA ACQUISITION

While the synthetically-generated state data above presents a dataset where anomalousness is indistinguishable by first-order statistics, experimental data does exhibit characteristic frequencies and data distributions that alter through time. Therefore, the proposed method is also applied to experimental ball bearing vibration test data provided by previous work done by the Case Western Reserve University (CWRU). The publically available Bearing Data Center (<http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>) supplies numerous experimental vibration datasets acquired from accelerometers on a motor housing where research intentionally seeded faults by electro-discharge machining (EDM) processes. These experiments use a 2 hp Reliance electric motor where the test conditions of the motor (load and angular speed) were monitored in conjunction with the housing accelerometer vibration data. Appendix B provides the conditions for all test cases. All faults ranged between 0.007 in and 0.021 in on the inner raceway, outer raceway, or rolling element separately varying in motor loads of 0, 1, and 2 hp. For consistent analysis, data comes from the same accelerometer location, due to stationary outer race fault responses being dependent on this location.

For all test data, a SKF 6205-2RS JEM bearing supports the motor shaft. This bearing has $J_B = 9$ rolling elements. Accelerometer vibration data is collected at a 12,000 Hz sampling rate and converted into the dimensionless g value, or ratio to the gravitational constant $g = 32.2 \text{ ft/s}^2$. Figure 3.12 illustrates example vibration signals where faults were seeded at a 0.021 in diameter on separate elements. Note that the characteristic bearing frequencies are clearly observable in these cases. For these signals, with the given inner ring rotating at $w_{IR} = 1750 \text{ rpm}$,

the calculated fault frequency values are: $FTF = 11.6$ Hz, $BPFI = 157.9$ Hz , $BPFO = 104.6$ Hz , and $BSF = 137.5$ Hz. In Figure 3.12 (b), $1/BPFO$ is equivalent to 0.01 seconds and $1/RPM$ is equal to 0.034 seconds. A vibrational impulse occurs at each $BPFO$. Note the changing amplitudes at each $BPFO$, as described in Section 2.2.2.

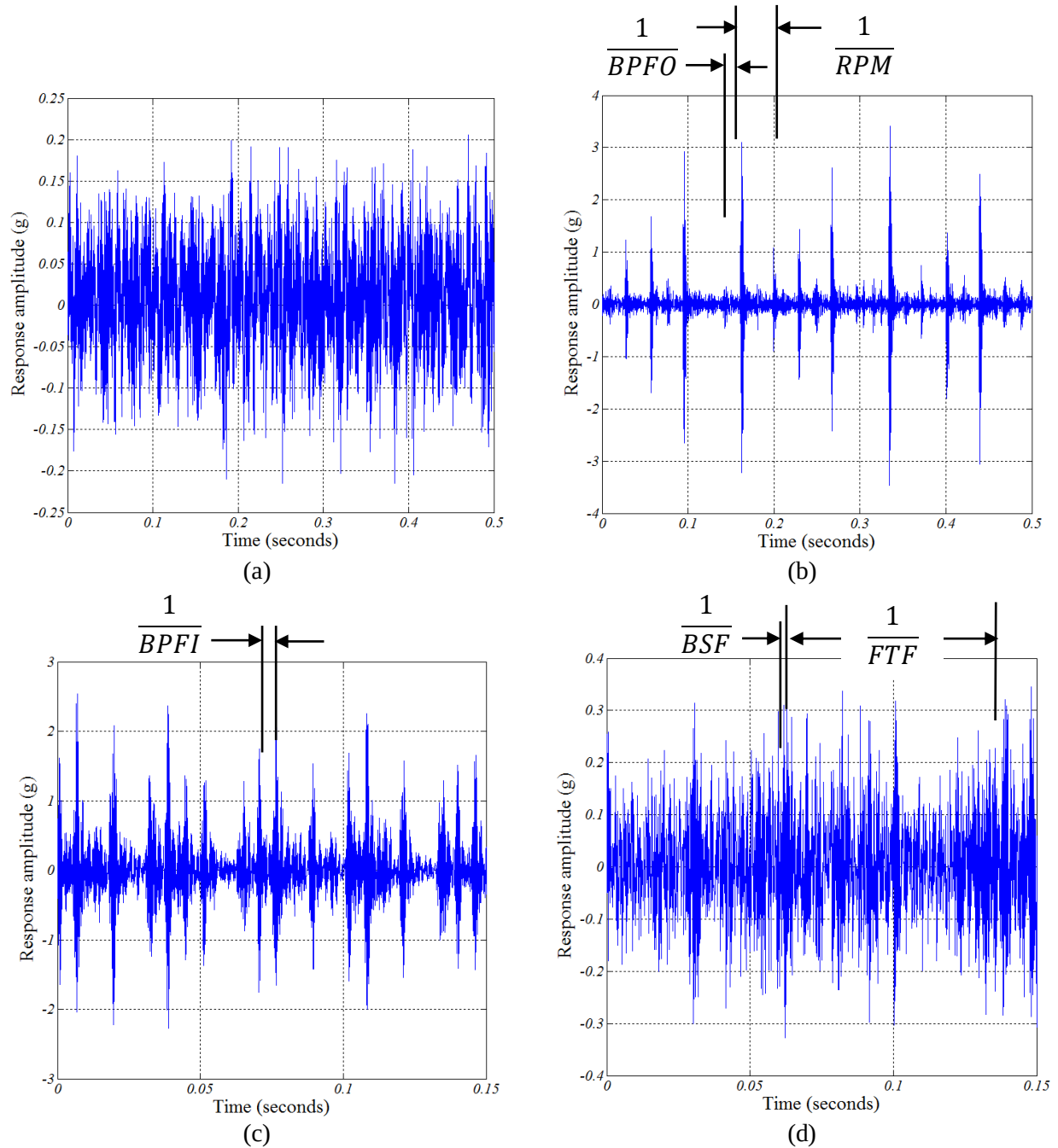


Figure 3.12. (a) Normal signal and (b) OR, (c) IR, and (d) RE signals.

Figure 3.12 (a) shows the response of a normal bearing. Note the relative noisy, low amplitude, high frequency signature. Figure 3.12 (b) through (d) demonstrate the fault-produced, higher amplitude shock responses within this noise. The overall signal distribution becomes less Gaussian, as more occurrences in the more extreme data values take place [110]. Rolling elements faults, seen in Figure 3.12 (d), produce vibratory amplitudes that are much lower than the impulses given off by inner or outer fault types. Since these signals correspond to a fixed outer race, the amplitudes for an outer race fault are sensitive to the position of the accelerometer with respect to this outer race fault, but will yield similar fault frequencies. In general, vibration frequencies indicate the source, while the amplitude indicates the severity [111]. In various situations, vibration analysis can also provide information about a bearing's rotational imbalance, misalignment, looseness, and resonance/critical speed situations [112].

3.4.2 ASSUMPTION VALIDATION

The distribution of the normal vibration data above, as well as the generalization of all normal bearing vibration signals, is shown below in Figure 3.13 (a). Since the variation in undamaged bearing vibration is the combination of separate and independent effects, the central limit theorem implies that this data is Gaussian, which is confirmed in practice [93]. Based on this assumption and the lack of evidence to prove that this data comes from a distribution that isn't normal, SAX partitioning of the time domain signal is appropriate in good bearings. A linear quantile plot will further validate this normality. However, damaged bearing signals display more extreme amplitude levels of vibration resembling more of what is statistically known as an alpha-stable distribution [110], which describe many experimental signals containing impulsive waveforms. Figure 3.13 (b) demonstrates this impulsive distribution for an outer race fault signature under a 0 hp motor load and a 0.021 in fault. Note the data does not follow a strict

normal probability density distribution, as shown in red. Further analysis shows that this change in distribution is just as severe in inner fault cases but less extreme in rolling element faults.

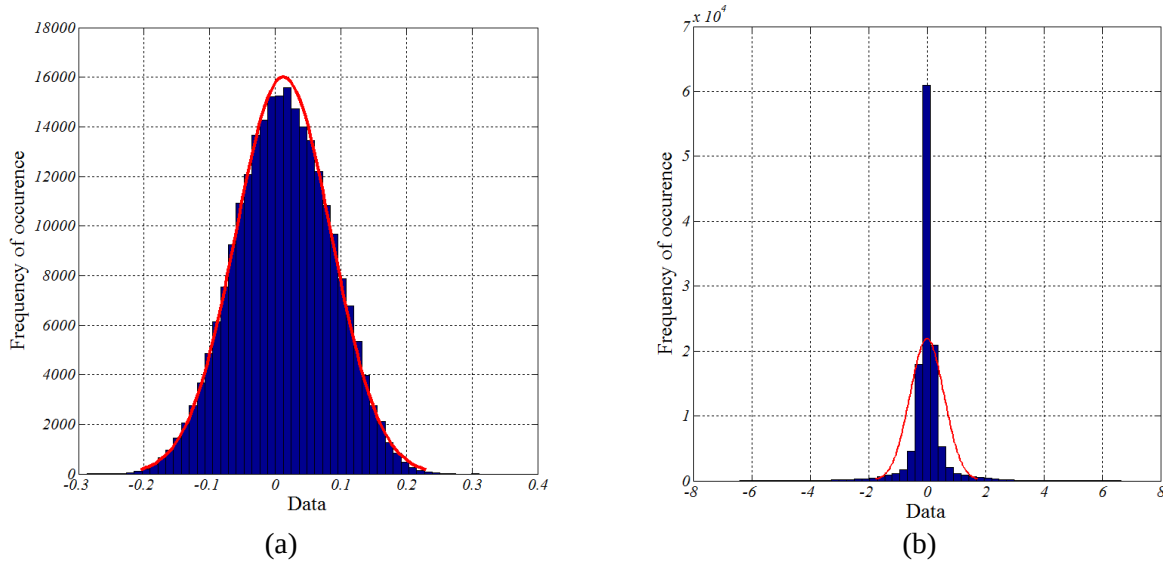


Figure 3.13. (a) Good and (b) faulty vibration time domain distributions.

SAX partitions data based on an assumed normal distribution. This assumption does not change between datasets, whether the signals are normal or faulty. Thus, SAX is not valid in highly faulty cases [47]. Any deviation between datasets will innately disturb the partitioning scheme, even if the transition probability between a reduced-state set does not change. These impulsive waveforms can result in too few or too many symbols representing the majority of data [113]. Considering these transition probabilities are the basis for anomaly detection, accurate diagnosis of faulty bearings in a HMM process must begin with a distribution that does not change between good and faulty bearing vibration. SAX also has the advantage in that this particular symbolic process does not *require* a strict normal distribution [103], which is the case of time-frequency domain analysis. Figure 3.14 below displays the data distribution of the time-frequency domain, or spectrogram image, for good and faulty bearing vibration given the same datasets as in Figure 3.13. The data maintains a relatively Gaussian form with slight skewness in both conditions, despite the severe defect impulses.

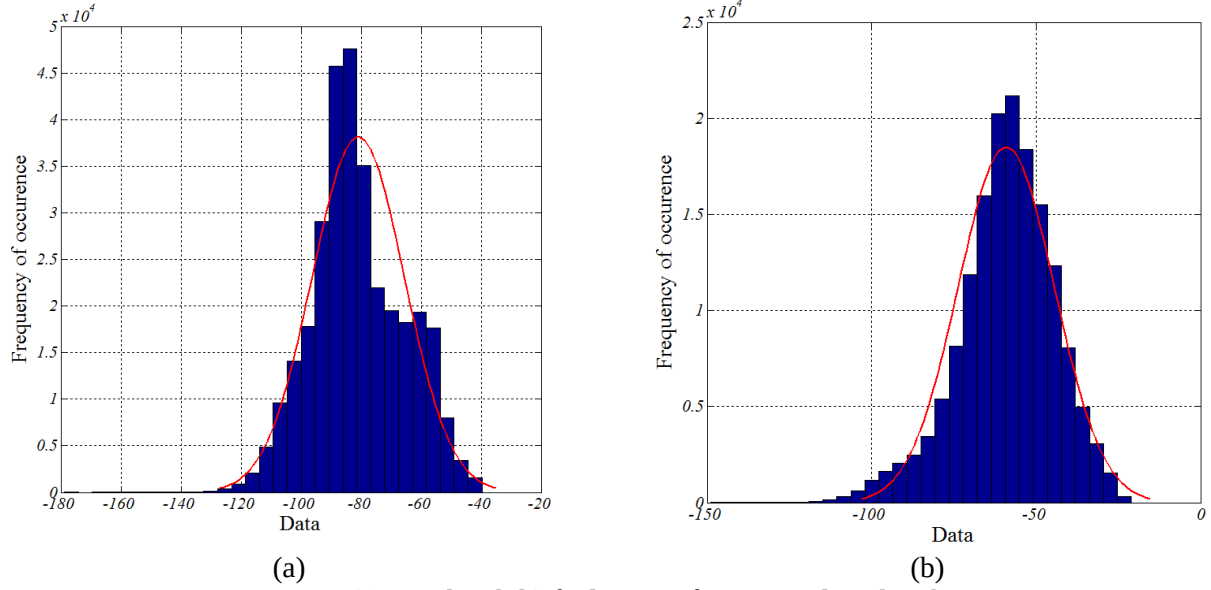


Figure 3.14. (a) Good and (b) faulty time-frequency data distributions.

The distribution of data does not change dramatically given a normal or faulty case scenario, allowing for consistent symbolic discretization and more accurate representation of state transition probabilities. The validation in this method's assumptions comes from the distribution of data post-partitioning. As shown in Figure 3.15 (a), after SAX partitioning using $N = 10$ and $\zeta = 1$, the number of each state is roughly equivalent in normal time domain data. However, in anomalous faulty conditions, such as in the outer race fault case in Figure 3.15 (b) the data represented by one state is significantly larger than other state. Note that states 2 and 9 represent little raw signal data. In the time-frequency domain of Figure 3.15 (c), the distribution of normal symbolic data is not quite equiprobable, indicating a non-strict normal distribution; however, this symbolic distribution between normal data and the outer race fault symbolic distribution, shown in in Figure 3.15 (d), provides that all states represent $10 \pm 8\%$ of raw data in the worst severity case. In time domain symbolic analysis, this variance is as high as 17%.

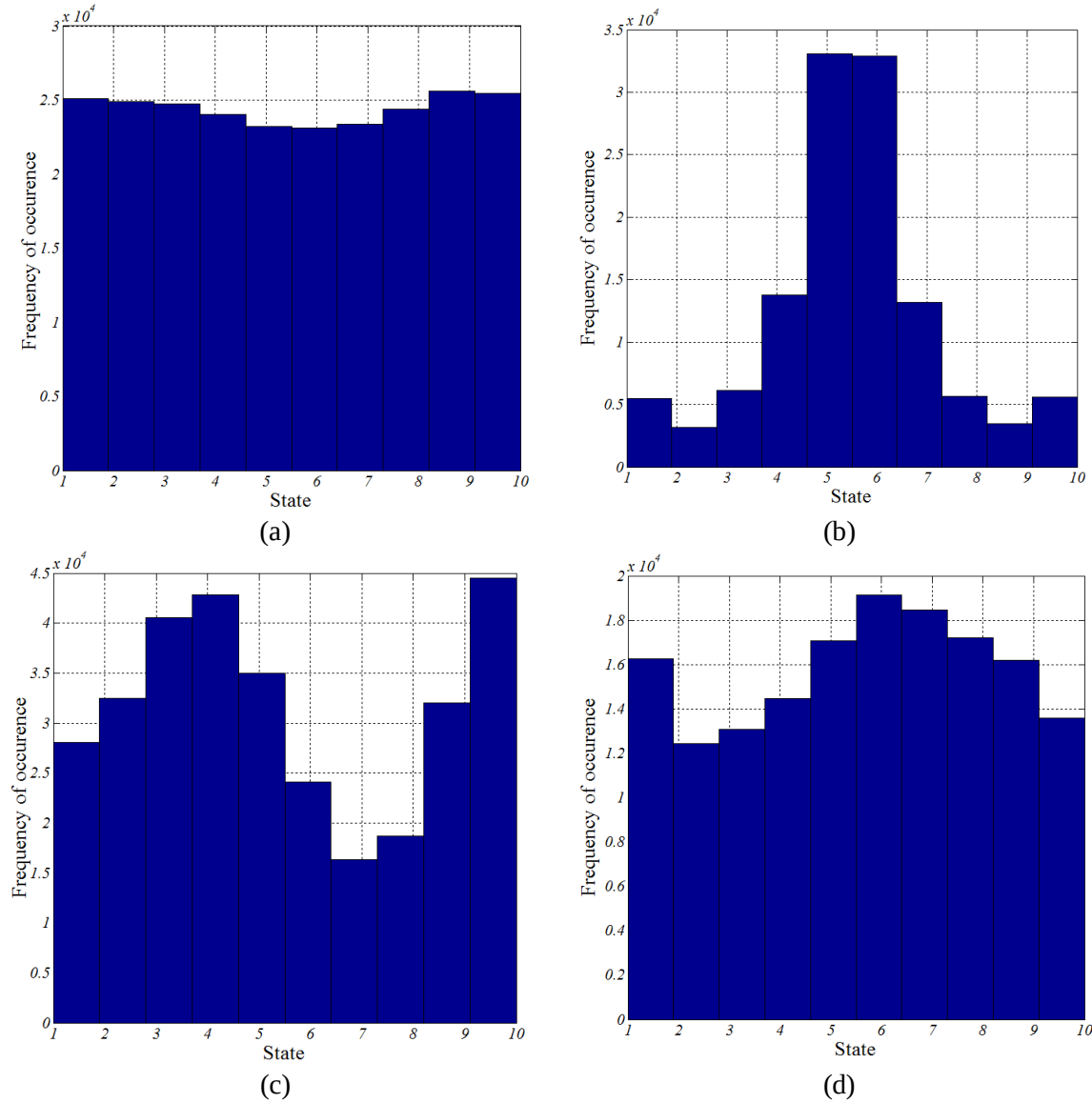


Figure 3.15. (a) Normal and (b) anomalous symbol distribution in the time domain. (c) Normal and (d) anomalous symbol distribution in the time-frequency domain.

While this experimental data is well-suited for fault detection performance comparison, the experimental aspect of condition monitoring cannot be analyzed through these datasets. CWRU experimental data contains preliminarily seeded faults and only small time segments (< 20 seconds) of data are available for each given faulty condition. Thus, the following numerical bearing vibration model is constructed to simulate propagating defects in bearings through time.

3.5 NUMERIC MULTIPLE DEGREE-OF-FREEDOM MODEL

Early detection of anomalies in roller element bearing data is critical in CBM applications. To analyze bearing vibrations where slowly developing faults are present, a numerical model of a bearing's response is constructed to simulate propagating defect vibration data. This multiple degree-of-freedom (MDOF) model is based on the work and validation of [114]. McFadden and Smith [31] first demonstrated models to simulate point defects in the early 1980's. Since then, more advanced bearing models, including the use of finite element (FE) analysis, are able to model non-linearity in the housings and rotor systems under different axial and thrust loads [115].

This specific MDOF application simply aims to model the bearing structure with a given forced vibration input from point defects. The structure simulates gradually propagating defects at different locations from the perspective of an accelerometer on the fixed outer race, where a propagating defect is modeled as a gap in material with a certain diameter. The calculation of the total vibrational response from the defects employs a quasi-static method, with one roller carrying the maximum load Q_m . Under moderate speeds and loading, this assumption is satisfactory. With the kinematic frequencies known from Section 2.2.2, only the specific characteristics of the excitation force are needed. The impact strength of a rolling element traversing over a localized defect depends of the speed and external load, thus both a static, given previously by Equation (2.1), and a dynamic component from the shock are present. Using Figure 3.16, conservation of mechanical energy between state 1 (before shock) and state 2 (after shock), is given by Equation (3.15):

$$\frac{1}{2}m_B V_{B1}^2 + \frac{1}{2}Iw_{B1}^2 + m_B g \left(\frac{\varphi_B}{2} \right) = \frac{1}{2}m_B V_{B2}^2 + \frac{1}{2}Iw_{B2}^2 + m_B g \left(\frac{\varphi_B}{2} \right) \cos \theta \quad (3.15)$$

where I_B is the ball mass moment of inertia ($I = 2/5 m_B (\varphi_B/2)^2$), V_B is the linear velocity in m/s and w_B is the angular velocity in rad/s ($V_B = (\varphi_B/2)w_B$).

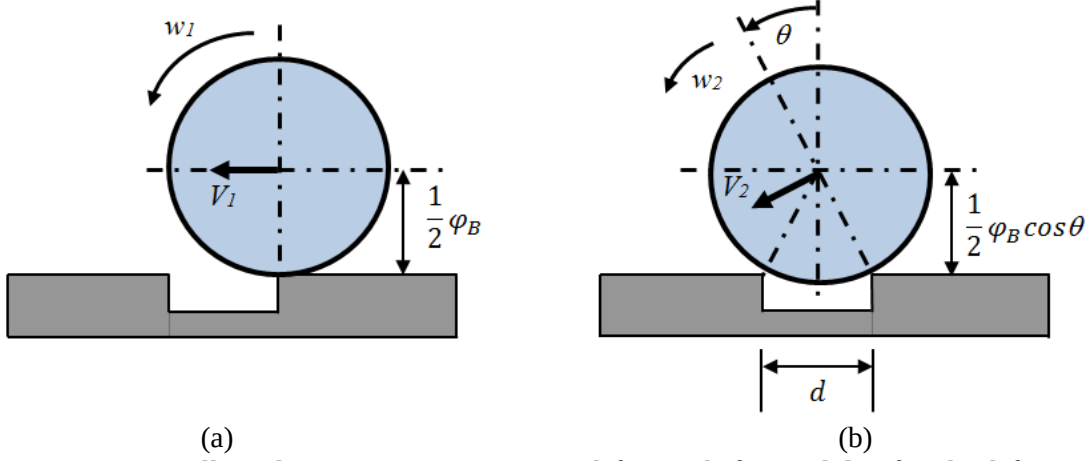


Figure 3.16. Rolling element passing over point defect (a) before and (b) after the defect.

Assuming small angle θ ($d/\varphi_B \cong \theta$), the change in velocity between states is:

$$\Delta V^2 = V_2^2 - V_1^2 = \frac{10}{28} g \varphi_B \left(\frac{d}{\varphi_B} \right)^2 \quad (3.16)$$

[116] presents evidence that steel impact force vary with the square of shock velocity. With this information, the full impact force F_T when a ball strikes a defect becomes:

$$F_T = Q_m [1 + \Gamma \Delta V^2] \quad (3.17)$$

where ΔV^2 comes from Equation (3.16), Q_m is the maximum static load from Equation (2.1), and Γ is the impacting coefficient based from the impact material and bearing geometry. These forces occur in pulses based on the geometric defect frequencies. Depending on the bearing type and material, these dynamic forces usually appear on the scale between $F_T = 800$ and 1200 N. A general vibration model is implemented based on Figure 3.17.

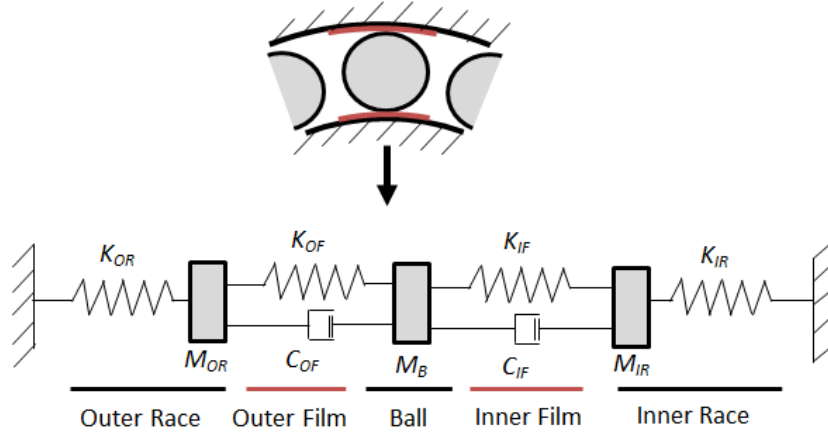


Figure 3.17. Rolling element bearing MDOF vibration model in the primary radial direction.

The behavior is assumed linear when the loading is moderate. Equation (3.18) below describes the bearing motion as a linear non-homogeneous second-order differential equation.

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K]\{y\} = \{F\} \quad (3.18)$$

where the components of the masses, stiffness's, and damping are found through:

$$[M] = \begin{bmatrix} M_{OR} & 0 & 0 \\ 0 & M_B & 0 \\ 0 & 0 & M_{IR} \end{bmatrix} \quad (3.19)$$

$$[K] = \begin{bmatrix} K_{OR} + K_{OF} & -K_{OF} & 0 \\ -K_{OF} & K_{OF} + K_{IF} & -K_{IF} \\ 0 & -K_{IF} & K_{IR} + K_{IF} \end{bmatrix} \quad (3.20)$$

$$[C] = \begin{bmatrix} C_{OF} & -C_{OF} & 0 \\ -C_{OF} & C_{OF} + C_{IF} & -C_{IF} \\ 0 & -C_{IF} & C_{IF} \end{bmatrix} \quad (3.21)$$

The displacement vector $\{y\}$ and the force vector $\{F\}$ from excitation shocks become:

$$\{y\} = \begin{Bmatrix} y_{OR} \\ y_B \\ y_{IR} \end{Bmatrix}, \quad \{F\} = \begin{Bmatrix} F_{OR} \\ F_B \\ F_{IR} \end{Bmatrix} \quad (3.22)$$

Depending on the fault type, the force F is 0 if no fault exists on that particular element; otherwise, F becomes F_T from Equation (3.17). A SIMULINK model (Appendix C) is developed to numerically solve Equation (3.18). All numerically-generated bearing vibration

responses were generated from parameters of a SKF 6205-2RS JEM deep groove ball bearing (25 mm bore, 52 mm outside diameter, 15 mm width, C3 clearance, 0 degree contact angle, and 9 balls with a pitch diameter to ball diameter of 4.9). Equivalent masses for M_{OR} and M_{IR} were calculated from Equation (3.23).

$$K = M * w_n^2 \quad (3.23)$$

where K, M , and w_n are the equivalent stiffness, equivalent mass, and natural frequency of the outer and inner races. Using Equation (3.24) [117], the natural frequencies in the second mode ($n = 0$ and $n = 1$ are rigid modes) for the outer and inner ring were 4102 Hz and 10593 Hz, respectively. FE simulations validate these values for the SKF 6205 bearing [114] with small relative errors of 1.25% and 6.4% between the calculated inner and outer ring flexural vibration natural frequencies, respectively. Thus, from Equation (3.23) above, the calculated equivalent stiffness's for the outer and inner race are $5.96(10)^7$ N/m and $2.48(10)^8$ N/m.

$$w_n = \frac{n[n^2 - 1]}{\sqrt{1 + n^2}} \sqrt{\frac{EI}{\Omega R^4}} \quad (3.24)$$

where w_n is the natural frequency of the ring in rad/s, n is the mode number, E is the modulus of longitudinal elasticity in N/m², R is the radius in meters, I is the moment of inertia about the cross-section of the ring in m⁴, and Ω is the mass per unit length in kg/m. The stiffness of the ball is left infinitely rigid in the model, as the equivalent ball stiffness is a minimum order of magnitude greater than any of the rings. Since other factors influence the vibratory response of a bearing in a faulty state, some assumptions are applied. The defect assumes a location in the maximum load radial direction ψ_m , where the defect impact force is greatest. While this assumption is not usually true from a given experimental accelerometer signal, this model simulates propagating faults and the angle from maximum load does not affect the validity.

Damping and stiffness values of the fluid film are calculated from Elastic-HydroDynamic theory (EHD) within the Hertzian contact zone between the rolling element and the raceways using short-width journal bearing theory. This thesis assumes equal damping and stiffness values in both fluid films ($K_{OF} = K_{IF} = K_F$ and $C_{IF} = C_{OF} = C_F$), although the minimum film thickness h_o occurs at the inner race. Further discussion of EHD is found in [118]. The calculation of the stiffness and damping matrices K and C are found in Appendix D.

Figure 3.18 (b) illustrates the dynamic response of this model under a preload of $Q_m = 100$ lb and an inner ring rotating at $w_i = 1750$ RPM and an outer race defect size of $d_{OR} = 0.021$ in. This signal provides a comparison by which to access the similarity of this model with CWRU experimental data under related parameters, shown in Figure 3.18 (a). Gaussian white noise is added to the signal based on the RMS value of the noise being equivalent to the RMS of the experimental simulated data subtracted from the RMS value of the CWRU data.

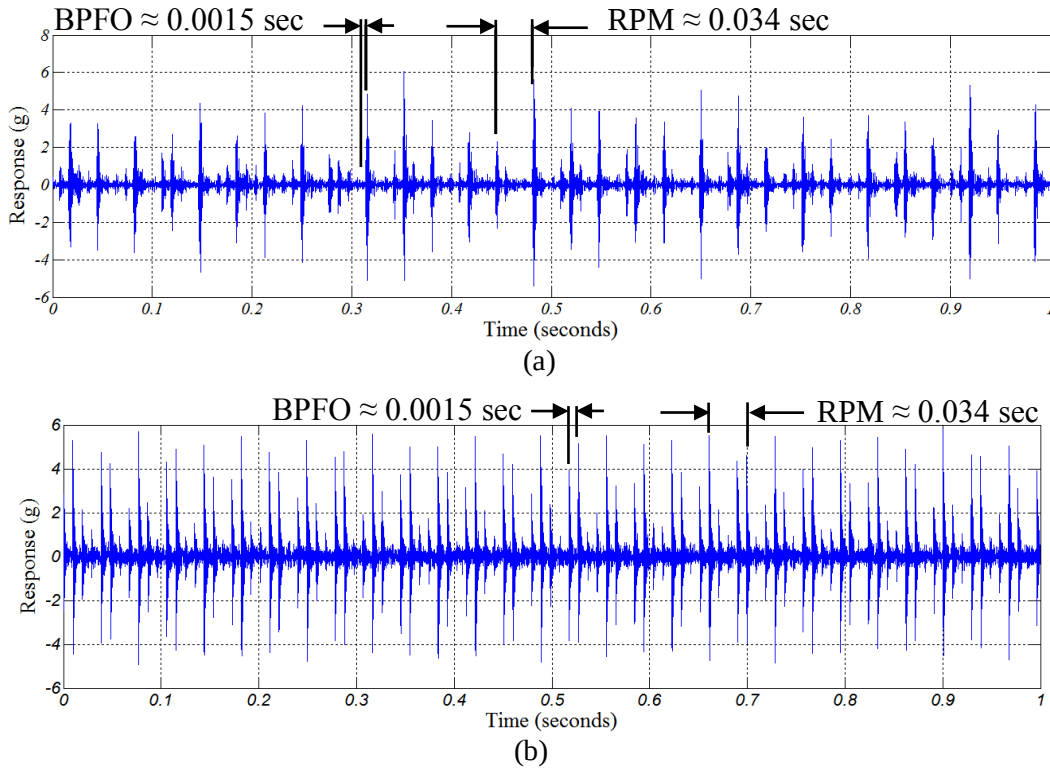


Figure 3.18. (a) CWRU experimental data. (b) Numerical model vibration response.

Table 3.2 through Table 3.4 provides validation of the numerical model with experimental data scalar indicators for each fault type. The simulated vibration signatures present a relative error of no more than 8.5% error in any time domain statistical scalar indicators. An amplitude correction is necessary due to the omission of simulated housings as well as other random non-correlating perturbations. The results were averaged over 5 simulations with 10 seconds of data each.

Table 3.2. Performance of numerical bearing vibration model with outer race fault.

	Experimental Results	Numerical MDOF Model	Relative Error (%)
<i>Peak</i>	1.69	1.83	8.3%
<i>RMS</i>	0.237	0.232	2.0%
<i>Crest Factor</i>	7.13	7.28	2.1%
<i>Kurtosis</i>	8.05	7.42	7.8%
<i>Impulse Factor</i>	10.39	10.8	3.9%

Table 3.3. Performance of numerical bearing vibration model with inner race fault.

	Experimental Results	Numerical MDOF Model	Relative Error (%)
<i>Peak</i>	3.62	3.32	8.1%
<i>RMS</i>	0.489	0.52	5.7%
<i>Crest Factor</i>	7.41	6.85	7.6%
<i>Kurtosis</i>	8.06	7.84	2.7%
<i>Impulse Factor</i>	11.12	10.25	7.8%

Table 3.4. Performance of numerical bearing vibration model with rolling element fault.

	Experimental Results	Numerical MDOF Model	Relative Error (%)
<i>Peak</i>	0.646	0.61	5.1%
<i>RMS</i>	0.107	0.101	5.6%
<i>Crest Factor</i>	6.02	5.91	1.8%
<i>Kurtosis</i>	3.3	3.04	7.8%
<i>Impulse Factor</i>	7.63	7.41	2.9%

This vibration model is also validated in the frequency domain. Figure 3.19 below provides the FFT by way of the envelop spectrum for the simulated data in Figure 3.18 (b). Figure 3.19 displays the BPFO = 656 Hz harmonics and the rotational frequency of $w_{IR} = 29$ Hz

as well as modulated sidebands, further demonstrating the qualitative and quantitative similarities.

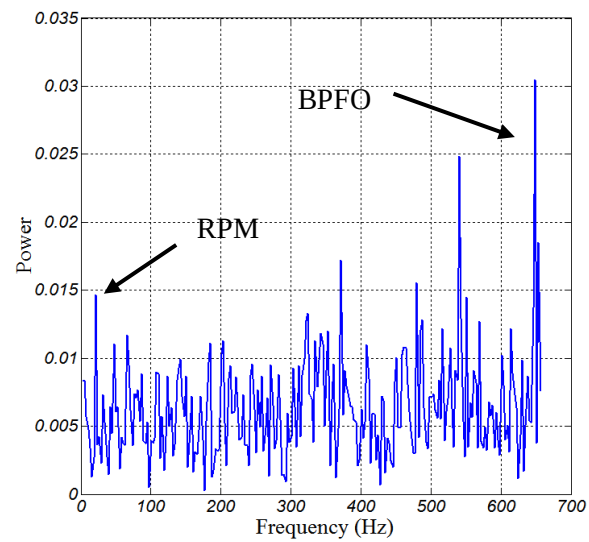


Figure 3.19. Spectrum of the simulated damaged bearing from Figure 3.18 (b).

4 COMPUTATIONAL PROCESS AND ANALYSIS

4.1 TRAINING AND TESTING PROCEDURES

Feature extraction of state data for anomaly detection is a classification-based procedure where anomalousness is determined through analysis of unknown data, called testing, by comparison of observations with known data, or trained data. Trained data encompasses the analysis of previously known observations in a method that is consistent with testing. In the analysis of all datasets for anomaly detection in this thesis, observations of normal data, or segments of normal data, are previously known. The TM resulting from SAX and subsequent HMM modeling of normal data is the basis for ‘normality’ in bearing vibration. Normality training is independent for each type of dataset. Testing involves the comparison of similar-length unknown segments of data in the same dataset to this ‘normal’ data in a windowing approach [5], where anomalousness is the scalar degree of difference between the normal and test TM’s for each test window. In short, the training method summarizes as:

1. Convert known normal time-series training raw into 2D spectrogram display P .
2. Perform SAX processing on spectrogram for symbolic data representation P_N .
3. Obtain TM through Bayes’ analysis of neighborhood transitions.
4. Parameterized TM into $n \times 1$ -sized PF_{normal} vector.

The testing method summarizes as:

1. Window unknown raw time-series data into β segments of length δ .
2. Convert each segment β of data into spectrogram display $P[\beta]$.
3. Perform SAX and state transform each $P[\beta]$ for symbolic/state discretization.
4. Obtain the TM $[\beta]$ for each $P_N'[\beta]$.
5. Parameterize each TM into $n \times 1$ -sized PF vector $PF_{test}[\beta]$.

6. Calculate anomalous score through l_1 distance $d(PF_{test}[\beta], PF_{normal}) \approx \eta$.
7. Specify anomaly threshold μ .
8. If η is above the μ , flag the window β as anomalous.

All analysis is done with MATLAB 2013a on a laptop computer running a 64 bit Windows 7 Service Pack 1 with an Intel® Core™ i7-2675QM CPU at 2.20 GHz and 6.00 GB of memory. The basic SAX code in both the 1D and 2D domain are found in Appendix E. The 2D state conversion and HMM modeling MATLAB code is found in Appendix F. Anomaly detection performance is based of true-positive rates (TPR) and false-positive rates (FPR) of detection. A TPR is a detection is which the algorithm successfully selects an anomalous conditions when, in fact, one does occur. A FPR, or Type II statistical error, is an algorithm's detection of anomaly, when the data detected as anomalous is, in fact, normal.

4.2 BISTOCHASTIC DATA PERFORMANCE

To test the proposed 2D SAX-based method, ground-truth normal state data generated based on Section 3.1 was processed to include known ground-truth anomalous data. A variety of non-overlapping anomaly sizes with a fixed window of observation were placed within normal data. That is, the tuning parameter for the performance metric is the length of anomalous data required for detection. The test set for this matrix occupied $J = 1(10)^7$ data points (with a y-axis resolution of 100 data points given 2D analysis), with anomalous data of varying lengths inserted at known times. These anomalies were adjusted to occupy shorter and shorter time-spans in the dataset, to test the HMM limit for anomaly discovery in a reduced anomalous state window.

For this study, the $1(10)^7$ point dataset was corrupted with 500 distinct spans of anomalous data in lengths of 1000, 900, 800, 700, 600, 500, 400, 300, 200, and 100 points. A

constant time segment of $\delta = 1000$ data point is the window of data length by which to determine segments of data normal or anomalous. Since this method assumes a spectrogram window length of 100 data points, the length of the two-dimensional datasets is in fact $1(10)^5$ and the anomaly window is 10 to maintain consistent analysis between 1D and 2D methods. The TPR and FPR were obtained by adjusting the threshold μ . The ground-truth performance characteristics of the anomaly detection method were collected and analyzed, and compiled into Receiver Operator Characteristic (ROC) curves shown below in Figure 4.1. The Non-Discriminate (ND) line is simply the 1:1 ratio of TPR to FPR.

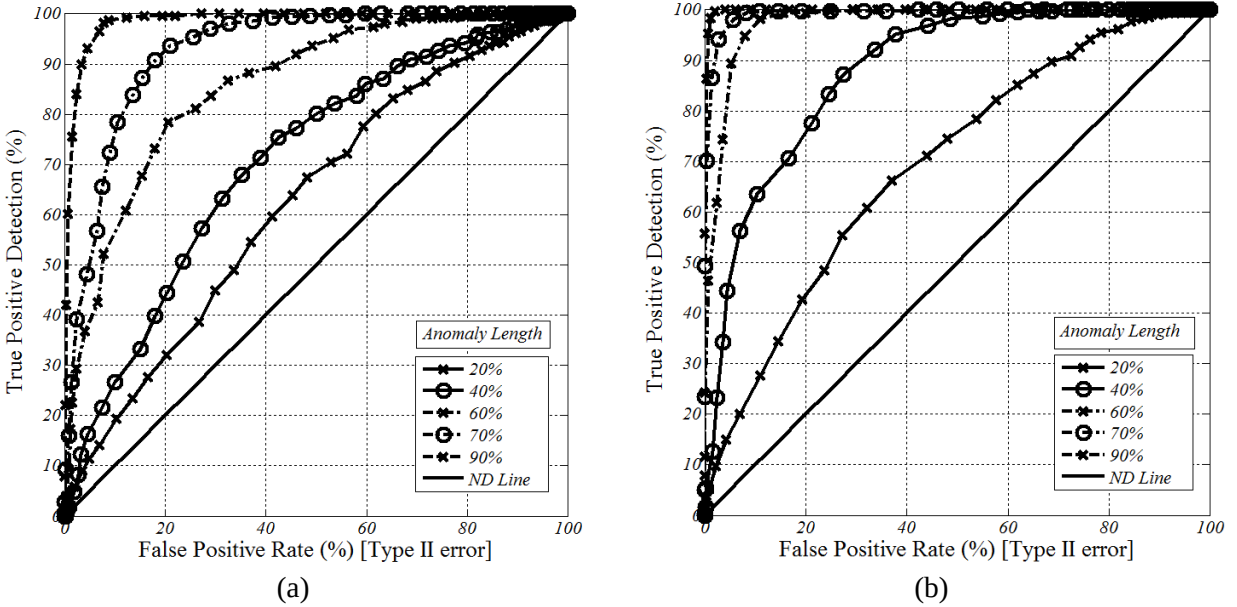


Figure 4.1. (a) 1D and (b) 2D bistochastic data performance.

From these results, 2D SAX outperforms simple 1D analysis. Even at only 60% of the data segment being anomalous (600 data points in the 1000 point window), the performance in 2D analysis is equivalent to 1D analysis where 90% of the segment is anomalous. In Figure 4.1 (b), it is any anomalous span of data greater than $O > 6N^2$ provides highly accurate detection rates – in this case, as high as 98% true-positive, and a $<1\%$ false-positive for a certain given threshold. However, this relationship will change for a different number of discrete states and

the amount of training data. This method still outperforms a random selector even as the anomaly size (< 1000 data points) is insufficient to fully populate a TM. Therefore, this method shows promise for sparse and noisy data. For a better illustration, Figure 4.2 demonstrates the basic results from this process using a $\delta = 1000$ data point window in 1D analysis. ‘Anomalous’ data windows comprise 50% of the entire dataset. Note the red zones indicate a detected anomaly at a specific threshold. There are both false-positive and missed detections, as presented below, resulting from anomalousness measures overlapping the selected threshold.

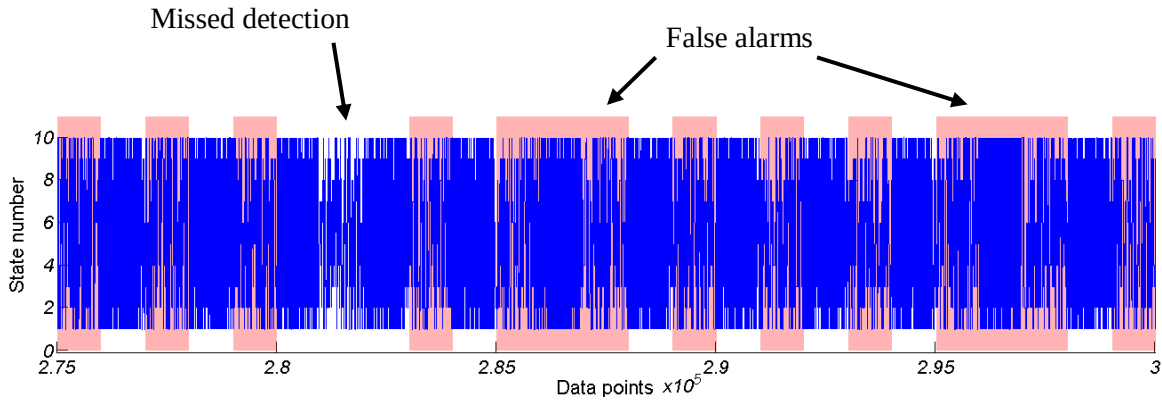


Figure 4.2. Example synthetic signal generated from bistochastic TMs.

Even though Section 3.3.1 provides that’s this synthetic state data is not a candidate for anomaly detection through first-order methods, a commonly used statistical method called the Nearest Neighbors (NN) approach is used for identifying portions of data which do not heuristically match a training set. NN uses a simple Euclidian distance comparison between two vectors of equal length. This approach is computationally easier to perform on reduced-dimensionality data having a discrete number of system states as opposed to continuous data. For reference, the Euclidian distance E between two vectors u and v is given by Equation (4.1):

$$E = \sum_{k=1}^K \sqrt{(u_k - v_k)^2} \quad (4.1)$$

In the NN approach, a system trains data by obtaining the distance measure E for each subset of training data against the rest of the training set. A portion of data which is farthest from its nearest neighbor is considered more ‘anomalous’. As the size of the training set decreases, this approach becomes less reliable; the probability of a test window existing within a training set diminishes. However, as the training set becomes too large, the probability of a similar or identical window of data existing within the training set grows. The NN method is applied to the time domain windowed data. Even in this common statistical method, a maximum TPR of 95.6% is achieved with a FPR of 16.2% failing to meet good detection criteria. The histogram range for this NN test’s anomalous measure results is shown in Figure 4.3, demonstrating a large overlap between normal and anomalous measures.

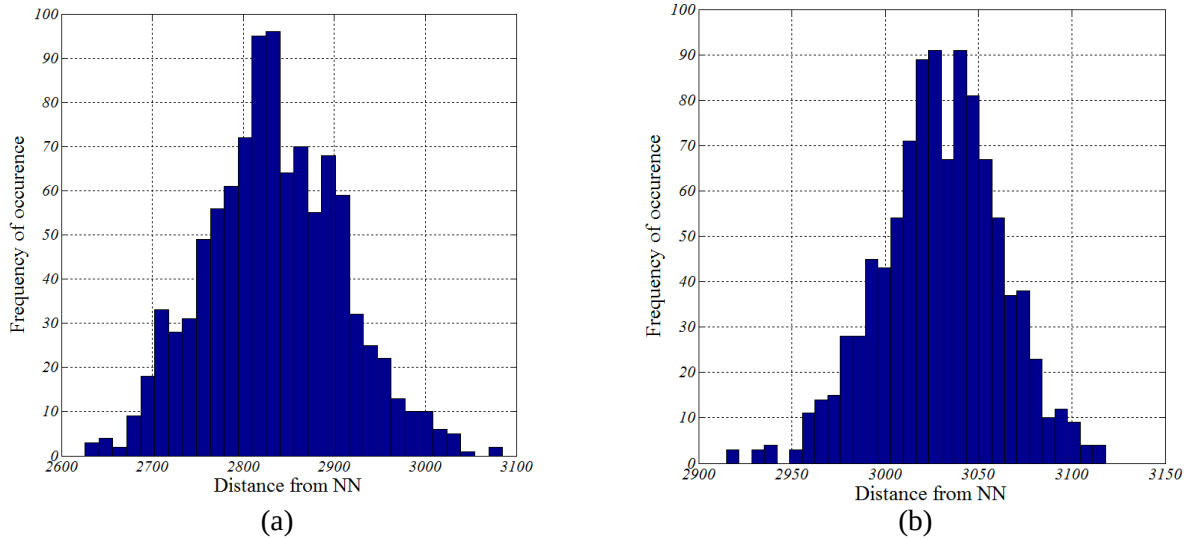


Figure 4.3. Euclidean NN for (a) normal data and (b) anomalous data.

Results were compiled from all methodologies applied to the data, and organized for comparison. Table 4.1 through Table 4.3 present a summary of training data types, anomaly sizes, performance metrics, and computation times for this synthetic data:

Table 4.1. Ground-truth bistochastic data, using state-based Nearest Neighbors.

# Training Points	# Test Points	% True Positive	% False Positive	Time to Compute
100,000	1,000,000	95.6%	16.2%	0.73s
10,000	1,000,000	91.7%	50.2%	0.072s
10,000	100,000	90.2%	59.4%	0.075s

Table 4.2. Ground-truth bistochastic data, using one-dimensional SAX.

# Training Points	# Test Points	% True Positive	% False Positive	Time to Compute
10,000	1,000,000	98.6%	8.3%	0.331s
1,000	1,000,000	92.1%	43.8%	0.330s
500	100,000	85.8%	73.2%	0.047s

Table 4.3. Ground-truth bistochastic data, using two-dimensional SAX.

# Training Points	# Test Points	% True Positive	% False Positive	Time to Compute
10,000	1,000,000	99.4%	7.6%	0.983s
1,000	1,000,000	95.6%	26.1%	0.975s
500	100,000	91.8%	33.1%	0.104s

The results in the above tables validate the advantage 2D symbolic modeling has in comparison against the popular NN and 1D symbolic modeling approaches. These results also demonstrate the reduced amount of training data needed for high accuracy of anomaly detection, especially in the both symbolic approaches. Note the increased performance with the increase in training data; however, the time for detection increases as well. Even though the above 2D simulated maps of low SNR data assumed a $\Delta = 100$ data point spectrogram window, this model still provides a greater amount of transitions per time period, thus populating the TM more quickly. The stopping rule in this analysis is when the spectrogram window becomes sufficiently large as to reduce the number of neighborhood transitions to fewer than the windowed time period (i.e., $\Delta > \text{frequency resolution}$).

4.3 EXPERIMENTAL DATA PERFORMANCE

Experimental machine vibration data behaves in a different manner than the synthetic state data above, particularly that specific and multiple fault frequencies exist. Since the CWRU datasets,

described in the previous chapter, are observed at a 12 kHz sampling rate, the 1000 data point anomaly window (or 0.0833 second time segment) that scans the data used for the synthetic data is also used for all the analysis in real-world data. Considering the lowest fault frequency occurs at a FTF = 11.6 Hz, this $\delta = 1000$ point window is applicable for this dataset's segmentation, as multiple revolutions of the bearing will occur within each time window. The following sections provide the analysis for 2D SAX performance on this real-world data and include the performance of different SAX parameters in an effort to investigate optimal performance of such a proposed algorithm.

4.3.1 PARAMETER SELECTION AND ANALYSIS

Figure 4.4 below demonstrates the performance of 2D SAX over traditional 1D symbolic analysis in a faulty outer race condition with 0.007 in defect. All anomalous vibration data is analyzed in conjunction with its relative normal dataset at the same motor load to establish ground-truth ROC curves. Note that here the 2D analysis requires $N = 6$ symbols for 100% anomaly detection and 0 false alarms. Even the use of $N = 5$ symbols provides the same high detection rate of 100% ground-truth accuracy with a 15% false-positive rate. On the other hand, the 1D analysis required minimum $N = 7$ symbols for perfect detection; whereas, the use of $N = 5$ symbols provides for a FPR of 25% to manage the same 100% detection accuracy.

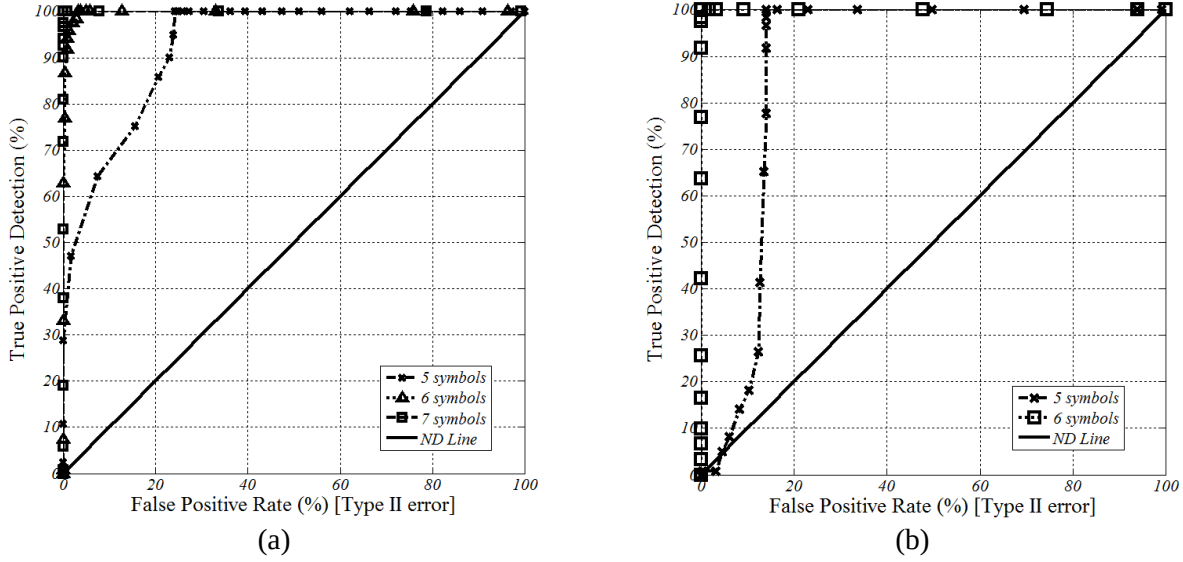
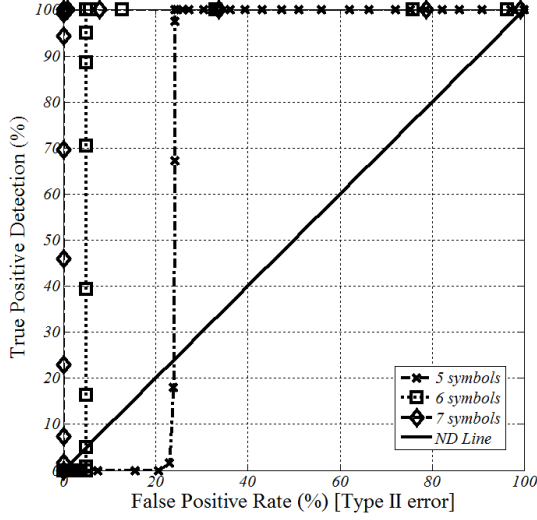
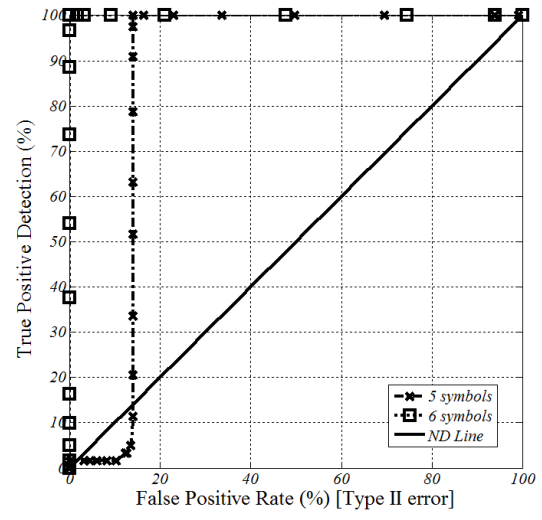


Figure 4.4. (a) 1D and (b) 2D alphabet size analysis for outer race fault.

The same alphabet size analysis is equally performed for inner race and rolling element faults, as seen in Figure 4.5 and Figure 4.6 below. The use of $N = 6$ symbols still results in a 100% true-positive detection with 0% false-positive rate in inner and rolling element faults. Again, 1D analysis requires $N = 7$ symbols for those detection rates. In all cases, increasing the alphabet size did not decrease performance up to a maximum tested $N = 20$ symbols. More interestingly, the performance of 2D SAX is much greater with a single localized fault presents itself on a rolling element. While 1D SAX demonstrates a severe lack in performance for this fault type (where the use of $N = 5$ symbols results in a minimum 23% FPR for a 100% TPR), as expected due to the low-amplitude random nature of rolling element faults, SAX in two dimensions maintains a high degree of performance. The similar performance between all datasets arises from the fact that this 2D anomaly detection technique is based on the detection of a propagating frequency, not a specific frequency.

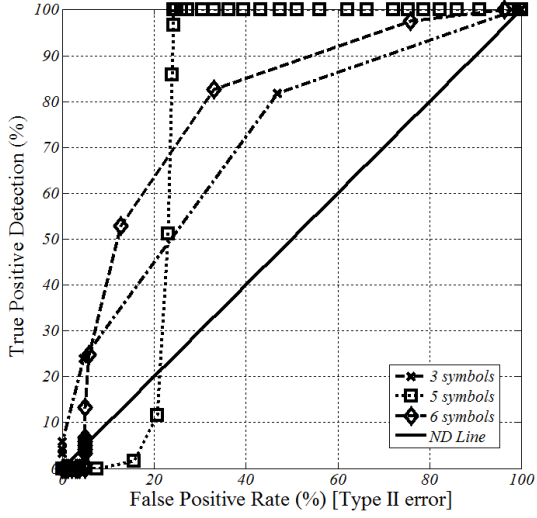


(a)

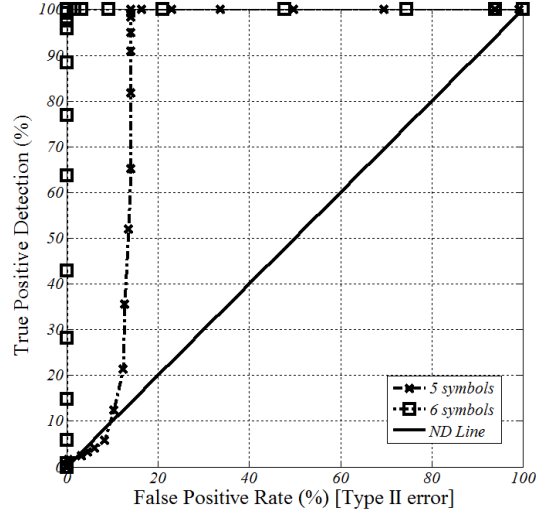


(b)

Figure 4.5. (a) 1D and (b) 2D alphabet size analysis for inner race fault.



(a)



(b)

Figure 4.6. (a) 1D and (b) 2D alphabet size analysis for rolling element fault.

2D SAX is also computationally-efficient in detecting anomalies. In all fault cases, analysis shows real-time capability as the amount of time needed for detection is less than the span of data that was analyzed, despite the alphabet size. Computational expense can also be reduced through spectrogram windowing. All previous analysis assumes a $\Delta = 100$ point spectrogram window for each $\delta = 1000$ data point segment. Table 4.4 presents anomaly detection results using a 200 and 50 point spectrogram window. The parameters for analysis

were $N = 6$, $\delta = 1000$ data points, and a 0.5 second training dataset on a dataset containing a 0.007 in OR fault.

Table 4.4. Ground-truth experimental vibration data, varying spectrogram window sizes.

Spec Window Length	% True Positive	% False Positive	Time to Compute
200 points	94.5%	7.2%	17.1s
100 points	99.7%	0.9%	28.7s
50 points	100%	0%	54.8s

Note the increased quicker execution time when under a spectrogram window of $\Delta = 200$ data points. However, this performance also decreases TPR by 5.2% and increases type II errors by 6.3%. Using a 50 data point spectrogram window length increases performance, but also increases computation time.

4.3.2 DISTRIBUTION ANALYSIS AND REGRESSION MODELING

This section provides analysis of the distribution of anomalous scores η . Figure 4.7 displays the distribution of anomalous scores for a bearing condition under motor load of 2 hp and a 0.007 in fault using the SAX and HMM transition modeling analysis using an alphabet size of $N = 6$.

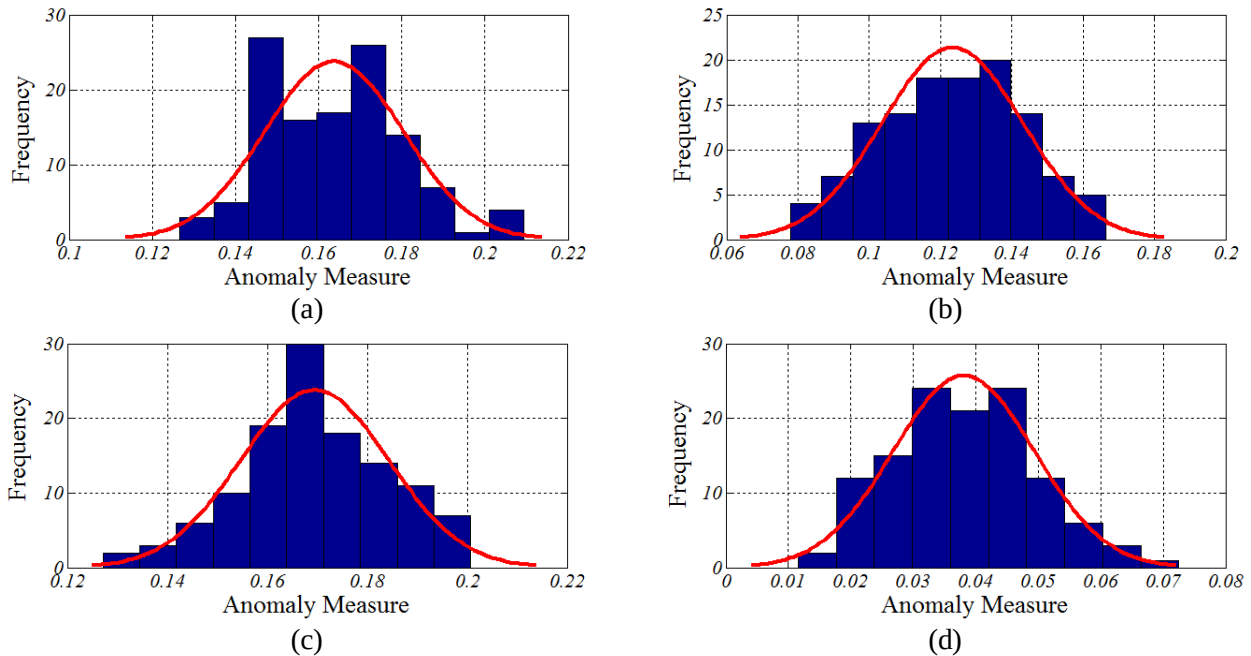


Figure 4.7. (a) OR, (b) IR, (c) RE fault and (d) normal anomaly distributions.

The anomalous scores shown above are based on a maximum normalized measure of 1. Also, note that anomalous measures for normal data are constant under all motor loads. Analysis of Variance (ANOVA) within this anomalous measurement data, revealed in Appendix G, demonstrates that the anomalous scores calculated from faulty bearing vibration have a statistically different mean from normal vibration data. 1D distribution analysis using $N = 7$, also shown in Appendix G, show 1D SAX is specifically incapable to distinguish rolling element faults from normal vibration data, even in high motor load scenarios.

This distribution analysis, when taken at different motor loads and fault severities, can serve as an agent for understanding the relationship between these anomalous detections and real-world vibration analysis. By extracting information from anomaly measures acquired at constant fault sizes, linear regression analysis provides an interpretation of 2D SAX-based anomalous measures and true fault severity through confidence-based anomaly detection. Using the anomaly measure distribution means and variances for each case, regression models are displayed in Figure 4.8 below. These figures illustrate the distributions of all anomaly measures at each fault severity with 90% and 95% confidence interval lines, shown in green and blue respectively. These models assumes anomalous scores are normally distributed, mutually-independent, and between each faulty signal, the ratio of maximum to minimum standard deviation is no more than 2. With a $s_{MAX}/s_{MIN} = 1.8$, this assumption is validated.

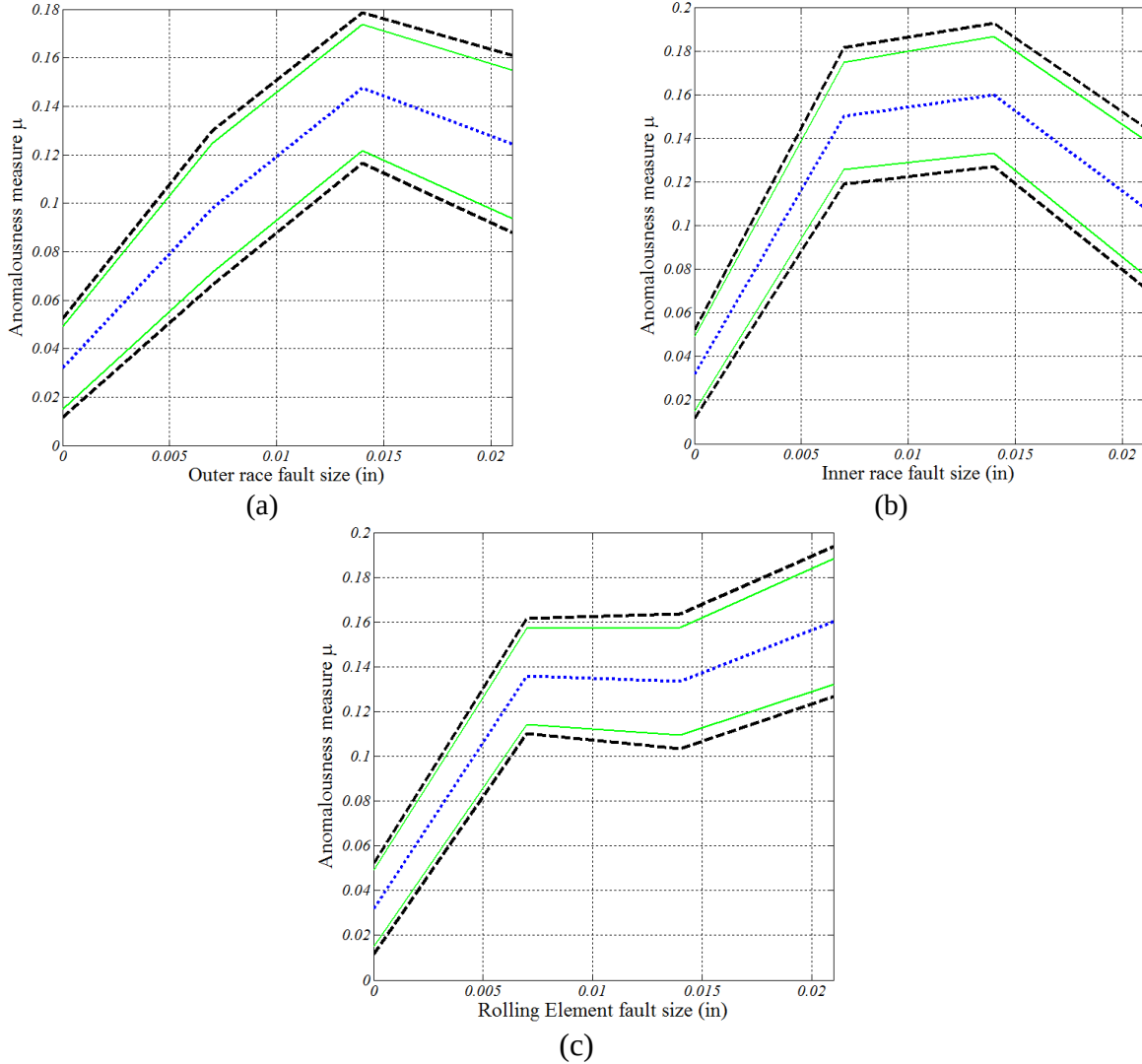


Figure 4.8. (a) OR (b) IR and (c) RE fault regression analysis.

Note the slightly increasing anomalousness deviations as the fault size increases, due to the increase in signal variability in more severe fault conditions. Also, at the higher fault size condition, the anomalous measures begin to actually decrease except in the case of a rolling element fault. This phenomenon is due to the fact that at the more severe conditions, higher occurrences of extreme states transitioning back to the same extreme state exist in a two-dimensional spectrogram. Table 4.5 below tabulates the means and 99%, 95%, and 90% confidence intervals for anomalous values that correlate to fault sizes of 0.007 in, 0.014 in, and 0.021 in.

Table 4.5. Anomalous measure given three different confidence intervals.

Fault Type	Anomaly Measure μ	Standard Deviation	99% C.I.		95% C.I.		90% C.I.	
			Lower	Upper	Lower	Upper	Lower	Upper
Normal	0.0321	0.01938	0.0053	0.0588	0.0117	0.0524	0.0150	0.0492
Outer Race	0.0980	0.0162	0.0562	0.1398	0.0663	0.1298	0.0714	0.1247
Outer Race	0.1477	0.0158	0.1068	0.1886	0.1166	0.1787	0.1216	0.1787
Outer Race	0.1242	0.0186	0.0762	0.1724	0.0877	0.1608	0.0936	0.1549
Inner Race	0.1502	0.0211	0.0959	0.2046	0.1189	0.1816	0.1256	0.1749
Inner Race	0.1598	0.0193	0.1100	0.2095	0.1270	0.2026	0.1331	0.1865
Inner Race	0.1086	0.0188	0.0601	0.1571	0.0718	0.1454	0.0777	0.1395
Ball	0.1357	0.0131	0.1018	0.1696	0.1100	0.1615	0.1141	0.1573
Ball	0.1335	0.0189	0.0847	0.1822	0.1034	0.1775	0.1094	0.1576
Ball	0.1602	0.0170	0.1162	0.2042	0.1269	0.1936	0.1322	0.1883

A more theoretical approach to classifying anomalousness at different fault sizes are presented in the following numerical vibration model results.

4.4 PROPAGATING FAULT MODEL PERFORMANCE

While the experimental data is a valid dataset for analysis of fully-developed fault types that do not change defect characteristics in time, the analysis of such data becomes limited for CBM applications. This section provides for the analysis of 2D SAX anomaly detection using accelerated propagating fault data, enabling comparison of 2D SAX with expectation in industrial environments.

Figure 4.9 (a) below displays an example of a progressing fault over time for an outer race propagating fault using the numerical bearing model described earlier. This signal is a simulated and accelerated case, where the fault size propagates far quicker than real-world scenarios. The fault propagates at a linear rate of $d(d_{OR})/dt = 4.0(10)^{-4}$ in/s for a total signal time of $t = 60$ seconds. The fault is set to being propagating at $t = 15$ seconds. Figure 4.9 (b) displays

the anomalousness measure through time as analyzed in this example vibration dataset. These scores are smoothed over a sliding window of 50 anomalous measures. Notice the increasing anomalous score with the increasing fault diameter. Upon the fault surpassing the noise threshold, the anomaly measure begins to increase exponentially, although small increases are visually noticeable before the noise threshold is broken by the fault signal, indicating the potential of this proposed method to detect faulty bearing data in environments low SNR.

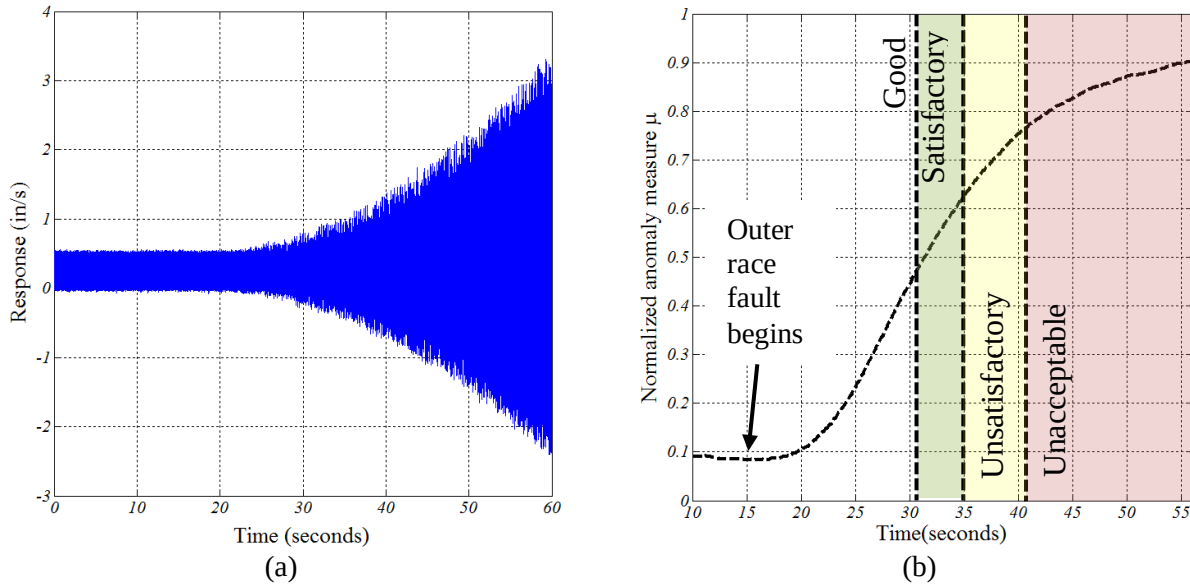


Figure 4.9. (a) Fault signal and (b) anomalous score with labeled ISO vibration standards.

Figure 4.9 (b) also correlates this continuous anomalous measure to ISO 10816 standards in Appendix A. By matching the theoretical anomalous scores and the known fault conditions, an accurate anomaly detection threshold can be selected based on vibrations standards. The RMS amplitude value of the velocity vibration signature is calculated in time and ISO thresholds are set at these specific times in Figure 4.9 (b). ISO standards assume RMS value in the velocity signature only change due to faulty signal data. At these calculated thresholds, theoretical anomalous measure can be selected based on industrial monitoring standards. Note in the above figure, these values correlate to 0.48, 0.61, and 0.78 which separate vibration data in ‘good’, ‘satisfactory’, ‘unsatisfactory’, and ‘unacceptable’ levels. In the above OR faulty case, where

unsatisfactory vibration is ~ 0.08 in/s or less based on ISO standards, an anomalous measure of ~ 0.6 is present. This correlates to a 0.008 in fault size. Note in this figure that the anomalous measures begin to increase rapidly far before industrial standards determine faults are unacceptable, indicating the early nature of detection of this method.

Figure 4.10 compares metrics by analyzing this numerically-generated signal in conjunction with time domain scalar quantities and the state-of-the-art 1D symbolic approach. The following curves demonstrate analysis on an anomalous signal spanning $t = 300$ seconds. Again, the outer race fault begins at $t = 15$ seconds, but increase at a reduced linear rate of $d(d_{OR})/dt = 5.0(10)^{-5}$ in/s. The 1D approach uses an alphabet size $N = 7$; whereas, the 2D SAX approach uses $N = 6$ to maintain high performance between each method. Each process is normalized to one, where one is the maximum recorded anomaly value specific to each individual analysis.

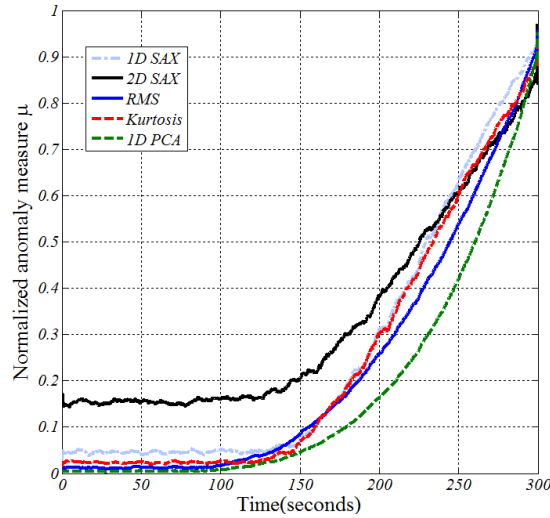


Figure 4.10. Anomalous scores through time in state-of-the-art methods.

Both a 1D and 2D covariance-based Principal Component Analysis (PCA) approach attempts to detect anomaly in these state-based propagating signals [119]. While 1D PCA testing shows reduced performance against other anomalous metrics in Figure 4.10, the 2D PCA

approach (not shown) is insensitive to spectrogram analysis, yielding no difference between normal and anomalous observations. This incompatibility in two dimensions is due to the lack of multiplicity within spectrogram data [120]. Similar to SAX, principal components are a reduced dimensionality of the original dataset by which to observe anomalousness. A FFT of the data at $t = 130$ seconds, or where the anomalousness gradient consistently increases in the 2D SAX approach, is shown in Figure 4.11. This figure illustrates the frequency domain response to that time segment consisting of 3000 data points. Even in the frequency domain, no concrete evidence exists to determine a faulty signal by the fault frequencies, where this figure exhibits a raised noise floor characteristic.

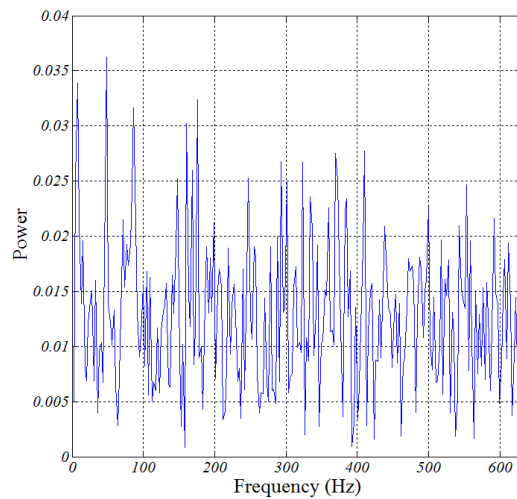


Figure 4.11. Frequency domain of propagating anomalous outer race signal.

Considering normality within each individual time segment is a relative measure between separate methods, an anomaly threshold that is consistent across all methods and consistent to each method's normality is necessary for accurate comparison. Table 4.6 below provides the results of anomaly detection for the 300 second OR propagating fault signal in Figure 4.10, where the first sign of detection is classified as a threshold of a 50% difference between the maximum anomaly value and the mean anomaly measure in the region of normality. This threshold ensures accurate comparison between individual methods. This process does not

include smoothing of measures, and detection is based on the first anomaly measure that surpasses this defined threshold.

Table 4.6. Condition monitoring comparison results.

	2D SAX	1D SAX	RMS	Kurtosis	1D PCA	Impulse
<i>First Anomalous Discovery</i>	212s	227s	238s	217s	219s	228s
<i>Time of Execution</i>	7.03s	1.37s	0.189s	0.808s	0.336s	0.327s

5 CRITICAL ANALYSES AND CONCLUSIONS

5.1 CRITICAL ANALYSIS OF BISTOCHASTIC DATA

Synthetically-generated symbolic data provides an environment where performance analysis of 1D and 2D HMM-based anomaly detection of state data is possible. While the explicit performance is only indicative of these specific datasets at $N = 10$, the relative performance between the two method will be consistent. As shown in Section 4.2, the 2D modeling outperforms simple 1D modeling, even in dataset cases where first-order statistic models cannot discriminate between normal and anomalous conditions. This increase in performance for this data is attributed to 2D HMM’s ability to more fully populate normal and test TM’s, as more transitions occur in a given time segment of data. For example, in a 0.1 seconds time window of vibration data (given a SAX window of $\delta = 1$ data point and $F_s = 12$ kHz sampling rate), TM’s model 1199 data samples. However, in the 2D approach (with a frequency resolution of a 100 and a similar SAX window), a TM will model 3600+ data point transitions per time segment.

In this synthetic reduced-dimensionality dataset, a nearest neighbor’s technique is also applicable and compared with 2D SAX. The performance of the NN approach demonstrates the inability to maintain characteristically ‘good’ detection rates, accounting for a 95% TPR in conjunction with a 16% FPR. Because this unsupervised approach to anomaly detection is purely data-driven, making no assumption about the generating distribution for the data, the distances between normal data observations and anomalous observations greatly overlap. The anomalous instances have too many close neighbors, resulting in missed anomalies and low false alarm rates. The computational complexity in the testing phase is also a significant challenge, since the distance measure of test instances is computed against all windows of training data to find the ‘nearest’ neighbor. This problem is highlighted in Table 4.2, where over 1×10^5 training

data points are necessary for the 95.6% true positive detection rate. Even though 2D symbolic modeling demonstrates three times greater computational expense, the detection rates are much greater and training data needed for high detection rates is much lower, even against its 1D symbolic competitor. Given similar length training sets, 2D SAX increases TPR by 0.8% and more importantly reduces FPR by 8% against the 1D SAX approach. Note that both these symbolic approaches also demonstrate the advantage that the length of training data does not significantly increase execution times.

This synthetic data necessarily sets a basis for test windowing performance, where this performance is dependent on the percentage of normal and anomalous data lengths in data segment windows. Given the performance ROC curves in Figure 4.1 (b), the $\delta = 1000$ data window is accurate for HMM anomaly detection, even without a fully populated TM, where 100% of anomalous signals are successfully detected when the entire data window is created by anomalous TMs. Reduced percentages, even as low as 60% of a time segment window generated by anomalous behavior, still provide high detection rates. While these symbolic datasets provide a metric for performance in state-based anomaly detection approaches, the specific data is not representative of true bearing vibration data. However, the capability in discerning windowing performance is limited in experimental data, as bearing vibration increases in faulty signals.

5.2 CRITICAL ANALYSIS OF EXPERIMENTAL FAULT DATA

The CWRU Bearing Data Center's datasets provide a source for experimental vibration tests where bearings undergo pre-defined faulty cases at different severities. The first application to note of this data is the ability for 2D SAX discretization and HMM transition modeling to

execute in real-time. Even with a spectrogram window of 50 data points (5% of the $\delta = 1000$ data point time segment), the total execution time totaled 49.2 seconds in over 300 seconds of vibration samples while being able to detect 100% of anomalies with 0 false positive readings in all datasets.

These experimental datasets also serve as an environment for symbolic parameter optimization in real data. Typically, an alphabet size $N = 10$ is a commonly used attribute for vibration discretization in various settings including bearing vibration partitioning [121], where little work has been done in selecting this parameter for performance. Analysis of CWRU bearing vibration data demonstrate an optimum alphabet size of $N = 6$. Higher magnitudes of the alphabet size did not reduce performance in 2D analysis; however, magnitudes greater than $N = 7$ reduced performance in discerning rolling element faults in 1D symbolic analysis. This alphabet size selection is specific to the high performance in these datasets. This high performance across all datasets in the 2D domain attributes to the independent nature between the alphabet size and the specific fault type in the time-frequency domain. Transition models take into account the neighborhood transitions, not the locations of fault frequencies in a spectrogram; thus, the numeric fault frequency (i.e., the specific fault location) does not change detection performance.

One interesting characteristic of 2D SAX is the previously mentioned ability to distinguish the low amplitude rolling element faults. From the figures in Section 4.3, the algorithm detects 100% of faulty rolling elements with 0 false alarms, for which the 1D SAX approach could only maintain this TPR with a 23% FPR. The main limitation in time-frequency partitioning is discretizing with too small an alphabet size. In these situations, anomalous data demonstrates more normality trends, especially in the case of inner race faults. However, with $N > 5$, this situation does not occur.

While this particular dataset is a good indicator of SAX performance, the analysis is limited by the fact that the data has already ‘progressed’ to a fully faulty state and faults do not change through time. Through regression analysis, a propagating fault can be simulated by displaying anomalousness information through increasingly known fault severity conditions. The regression analyses indicate the increasing anomalousness measures of bearing faults in earlier propagating fault sizes along with increasing variance, as expected. Thus, the confidence intervals will expand at larger fault sizes. At higher fault sizes, the gradient of the anomalous measurements decreases, especially between 0.021 in and 0.028 in fault sizes in IR and OR faults. This indicates another limitation in 2D symbolic dynamics: a maximum anomalousness measure does exist in experimental data. However, there does not exist the limitation that fault must not be too severe, which is the case in many time domain indicators (such as Kurtosis).

Unfortunately, through the distribution analysis of these anomaly measures, the ability to diagnosis fault types is not possible under this proposed algorithm. It is noted that this symbolic approach, whether in 1D or 2D, is an anomaly detection approach to bearing fault detection, and thus the ability to distinguish between faults is not applicable. By detecting ‘anomaly’, this process forgoes the ability to distinguish specific fault scenarios. Human interaction is needed upon detection for accurate diagnosis.

5.3 CRITICAL ANALYSIS OF PROPAGATING DEFECT DATA

The propagating bearing fault model attempts to model the conditioning monitoring aspect of bearing health. This analysis is critical in that statistical anomaly detection is an inverse and forward analysis problem. The forward problem aims to identify patterns in the dynamics and track statistical changes in time. However, this propagating fault vibration model of a bearing

aims to extract information from the observed time series data, or the inverse problem of anomaly detection. This type of analysis is seen in Figure 4.9. Using the ISO 10816 scale for acceptable vibration severity, ‘normal’ threshold values for anomalousness are made available. While theoretical, this model does give some insight into the correlation between normal operation and anomalousness measurements. Particularly that at a fault size of 0.008 inches, where anomalousness equates to 0.6 out of 1, current standards determine these vibration levels to be unsatisfactory. However, experimental data results in the previous section display a maximum anomalousness value of about 0.2, thus confirming the need for more experimental validation of condition monitoring through this proposed process. In all, this figure’s results suggest real-world thresholds for ‘unacceptable’ bearings may potentially be lowered in experimental data, thus reducing the time needed in the detection of faulty bearings.

The numerically-generated propagating fault data also serves as a performance metric compared with other commonly used methods of fault detection. Using the defined threshold of 50% of the difference between the maximum and average normal anomalous measures, 2D SAX detects an ‘anomalous’ bearing at a lower fault size (i.e., earlier in time), thus confirming the high performance in machine monitoring environments. While this simulate data’s faulty nature propagates far quicker than true experimental, this reduction in time can be days in advance in truly industrious environments. 2D SAX also detects anomalous and faulty bearing vibration data earlier in time than current on-line monitoring methods, including commonly used frequency methods and the statistical PCA method. Since 2D HMM detects extraneous fault frequencies that can exist under the noise threshold, this proposed architecture detects anomalies before other methods that require higher amplitudes, such as RMS and impulse factor values.

5.4 RESEARCH BENCHMARK AND CONCLUSIONS

This thesis provides the foundation for 2D machine learning-based processing in rolling element bearing anomaly detection. The basic approach involves symbolic dynamics for spectrogram image statistical modeling. While symbolic dynamics is typically a 1D approach to anomaly detection and not necessarily a novel approach to machine monitoring, the application and extension of 2D symbolic modeling in this area have been neglected in recent research endeavors, even though these 2D techniques of anomaly detection are extensively employed, such as in the case of target detection. This thesis' proposed method thus fills research gaps, providing a framework for further analysis in 2D analysis of bearing fault data.

This thesis presents these results of 2D bearing vibration analysis particularly in the symbolic domain. By actively selecting discretization parameters, a benchmark for experimentally optimizing partitioning parameters is laid for bearing vibration data. By comparing 1D and 2D symbolic dynamics of bearing data, this thesis provides a comparison metric by which to demonstrate the advantages 2D SAX analysis presents in bearing anomaly detection over traditional time domain signal analysis. This proposed method further demonstrates novelty in that the 2D HMM modeling of vibration data allows for machine learning-based anomaly detection, without the use of human expertise for anomalous signal matching, in spectrograms. From condition monitoring analysis of numeric ball bearing vibration data, this method is a real-time approach which can be easily implemented on inexpensive platforms. While this thesis presents many unique ideas in the area of bearing anomaly detection, this analysis lays a foundation in the area of anomaly detection and presents the necessity to expand on the subject is presented in the following section.

5.5 SHORT-TERM AND LONG-TERM FUTURE WORK

While this method is a viable option for machine fault detection in industrious environments, some improvements are needed in both short and long term research. Since the method detects anomalous frequencies, limitations exist in that anomalousness is the indication of an unidentified fault frequency. More analysis is needed for different anomalous conditions, such as shaft misalignments and structure looseness. This analysis may include the use of other intelligent anomaly discovery techniques, such as hierarchal statistic models as opposed to HMM, to classify the changing patterns of such faulty conditions. The main advantage in adapting other anomaly discovery environments is the ability to model frequency location dependency and discern fault types. While the SAX alphabet size parameter N is selected based on data-driven performance, a simulated design of experiments is necessary for full optimization of all SAX parameters, such as window length and overlap value. This thesis simply assumes changing spectrogram window sizes, and not necessarily the SAX window size will change performance, and thus the ability of one symbol to represent larger or smaller areas of vibration in the spectrogram image may alter the performance of the algorithm.

In all, more extensive experimental work is necessary, especially in the aspects of condition monitoring, where slow-time scale pattern recognition is of critical nature. This includes work incorporating run-to-failure conditions, where analysis of truly propagating failure conditions can be monitored. A prognostic correlation between given anomalous scoring and remaining life measures is the necessary end result of this analysis. Currently, an experimental bearing rig is being constructed for such work, as seen in Appendix H. Despite the extensive work needed to further validate 2D SAX to bearing anomaly detection to real-world condition

monitoring, this thesis provides the fundamentals and preliminary research that demonstrates the effectiveness such as method has in the industry field.

REFERENCES

- [1] F. Edgeworth, "On discordant observations," *Philosophical Magazine*, vol. 23, no. 5, pp. 364-375, 1887.
- [2] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, p. 15, 2009.
- [3] H.-F. Ng, "Automatic thresholding for defect detection," *Pattern Recognition Letters*, vol. 27, no. 14, pp. 1644-1649, 2006.
- [4] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Information Sciences*, vol. 177, pp. 3799-3821, 2007.
- [5] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448-3470, 2007.
- [6] I. Steinwart, D. Hush and C. Scovel, "A classification framework for anomaly detection," *Journal of Machine Learning Research*, vol. 6, pp. 211-232, 2005.
- [7] S. Karpagachelvi, M. Arthanari and M. Sivakumar, "ECG feature extraction techniques - A survey approach," *International Journal of Computer Science and Information Security*, vol. 8, no. 1, pp. 76-80, 2010.
- [8] M. Das, "Spatio-temporal anomaly detection," Columbus, OH, USA, 2009.
- [9] S. Ghosh and D. Reilly, "Credit card fraud detection with a neural-network," in *27th Hawaii International Conference on System Sciences*, Grand Wailea, Maui, HI, USA, 1994.
- [10] O. Boiman, E. Shechtman and M. Irani, "In defense of nearest-neighbor based image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, 2008.
- [11] D. Barbara, J. Couto, S. Jajodia and N. Wu, "Detecting novel network intrusions using Bayes estimators," in *1st SIAM International Conference on Data Mining*, Chicago, IL, USA, 2001.
- [12] F. Anscombe and I. Guttman, "Rejection of outliers," *Technometrics*, vol. 2, no. 2, pp. 123-147, 1960.
- [13] X. Hu, R. Subbu, P. Bonissone, H. Qiu and N. Iyer, "Multivariate anomaly detection in real-world industrial systems," in *IEEE International Joint Joint Conference on Neural Networks. (IEEE World Congress on Computational Intelligence)*, Hong Kong, China, 2008.
- [14] E. Keller and A. Ray, "Real-time health monitoring of mechanical structures," *Strutural Health Monitoring*, vol. 2, no. 3, pp. 191-203, 2003.
- [15] A. Kumar, A. Srivastava, A. Banerjee and A. Goel, "Perfromance based anomaly detection analysis of a gas turbine engine by artifical neural network approach," in *Annual Conference of Prognostics and Health Management Society*, Minneapolis, MN, USA, 2012.
- [16] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining

- processes," *International Journal of Production Research*, vol. 43, no. 6, pp. 1275-1293, 2005.
- [17] S. McArthur, C. Booth, J. McDonald and I. McFayden, "An agent-based anomaly detection architecture for condition monitoring," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1675-1682, 2005.
 - [18] C. Harris, A. Piersol and T. Paez, *Shock and vibration handbook*, 6 ed., McGraw Hill, 2009.
 - [19] W. Zhang, Q. Yang and Y. Geng, "A survey of anomaly detection methods in networks," in *International Symposium on Computer Network and Multimedia Technology*, Wuhan, China, 2009.
 - [20] S. Ericsson, N. Grip, E. Johansson, L.-E. Persson, R. Sjoberg and J.-O. Stromberg, "Towards automatic detection of local bearing defects in rotating machines," *Mechanical Systems and Signal Processing*, vol. 19, no. 3, pp. 509-535, 2005.
 - [21] B. J. Hamrock and W. J. Anderson, "Rolling-element bearings," 1983.
 - [22] J. E. Shingley and C. R. Mischke, *Mechanical engineering design*, 6 ed., McGraw-Hill, 2001.
 - [23] F. Sadeghi, B. Jalalahmadi, T. S. Slack, N. Raje and N. K. Arakere, "A review of rolling contact fatigue," *Journal of Tribology*, vol. 131, p. 041403, 2009.
 - [24] C. Sunnersjo, "Rolling element bearings - the effects of geometric imperfections and wear," *Journal of Sound and Vibration*, vol. 98, no. 4, pp. 455-474, 1985.
 - [25] I. Y. Onel, K. B. Dalci and I. Senol, "Detection of bearing defects in three-phase induction motors using Park's transform and radial basis function neural networks," *Sadhana*, vol. 31, no. 3, pp. 235-244, 2006.
 - [26] A. Voskamp, R. Osterlund, P. Becker and O. Vingsbo, "Gradual changes in residual stress and microstructure during contact fatigue in ball bearings," *Metals Technology*, vol. 7, pp. 14-21, 1980.
 - [27] S. Lacey, "An overview of bearing vibration analysis," *Maintenance & Asset Management*, vol. 23, no. 6, pp. 32-42, 2008.
 - [28] S. Lacey, "Using vibration analysis to detect early failure of bearings," 2007.
 - [29] H. J. Jung, K. H. Baek, S. Hidaka, D. Song, S. B. Kim and T. H. Sung, "Design and optimization of secondary shock type piezoelectric system," in *International Symposium on Applications of Ferroelectrics*, Aveiro, Portugal, 2012.
 - [30] A. Bhende, G. Awari and S. Untawale, "Assessment of bearing fault detection using vibration signal analysis," *Technical & Non-Technical Journal of Visual Soft Research Development*, vol. 2, no. 5, pp. 249-261, 2011.
 - [31] P. McFadden and J. Smith, "Vibration monitoring of rolling element bearings by the high-frequency resonance technique - a review," *Tribology International*, vol. 17, no. 1, pp. 3-10, 1984.
 - [32] J. Wensing, "On the dynamics of ball bearings," Enschede, The Netherlands, 1998.
 - [33] Z. Yanbing, L. Yibing, A. Hongwen and C. Liming, "Higher order spectral analysis for vibration signals of the large steam turbine in slow-down process," in *International Conference on Intelligent Control and Information Processing*, Harbin, China, 2011.

- [34] H. Akturk, M. Uneeb and R. Gohar, "The effects of number of balls and preload on vibrations associated with ball bearings," *Journal of Tribology*, vol. 119, pp. 747-753, 1997.
- [35] L. Tatjana, R. Mileta and M. Radivoje, "Load distribution in rolling bearings," *FME Transactions*, vol. 36, no. 4, pp. 189-196, 2008.
- [36] B. Al-Najjar, "Improved effectiveness of vibration monitoring of rolling bearings in paper mills," *Journal of Engineering Tribology*, vol. 212, no. 2, pp. 111-120, 1998.
- [37] P. Staff, "Plant Engineering & Maintenance," 2011. [Online]. Available: [http://www.pem-mag.com/News/Global-machine-condition-monitoring-market-to-reach-US\\$2.1B-by-2015.html](http://www.pem-mag.com/News/Global-machine-condition-monitoring-market-to-reach-US$2.1B-by-2015.html). [Accessed 22 2 2013].
- [38] Y.-H. Kim, A. C. Tan, J. Mathew and B.-S. Yang, "Condition monitoring of low speed bearing: a comparative study of the ultrasound technique versus vibration measurements," *Engineering Asset Management*, pp. 182-191, 2006.
- [39] J. R. Clifton and N. J. Carino, "Nondestructive evaluation methods for quality acceptance of installed building materials," *Journal of Research of the National Bureau of Standards*, vol. 87, no. 5, pp. 407-438, 1982.
- [40] R. Draney, "High temperature sensor for bearing health monitoring," in *IEEE Aerospace Conference*, Big Sky, MT, USA, 2008.
- [41] R. Yan and R. X. Gao, "Harmonic wavelet-based data filtering for enhanced machine defect identification," *Journal of Sound and Vibration*, vol. 329, no. 15, pp. 3203-3217, 2010.
- [42] T. Rathbone, "Vibration tolerance," *Power Plant Engineering*, vol. 43, p. 721, 1939.
- [43] T. Doguer and J. Strackeljan, "Vibration analysis using time domain methods for the detection of small roller bearing defects," in *8th International Conference on Vibration in Rotating Machines*, Swansea, UK, 2009.
- [44] A. Roque, T. Silva, J. Calado and J. Dias, "An approach to fault diagnosis of rolling bearings," *WSEAS Transactions on Systems and Control*, vol. 4, no. 4, pp. 188-197, 2009.
- [45] C. Tan and Y. Okada, "Detection of rolling-element bearing signal corrupted by noise of similar frequency using adaptive noise cancellation," in *International Conference on Information, Communications and Signal Processing*, Singapore, Singapore, 1997.
- [46] J. Crutchfield and N. Packard, "Symbolic dynamics of one-dimensional maps: entropies, finite precision, and noise," *International Journal of Theoretical Physics*, vol. 21, no. 6/7, pp. 433-466, 1982.
- [47] F. de Lorenzo and M. Calabro, "Kurtosis: a statistical approach to identifying defect in rolling bearings".
- [48] R. Heng and M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition," *Applied Acoustics*, vol. 53, no. 1-3, pp. 211-226, 1998.
- [49] P. Stoica and R. Moses, *Spectral analysis of signals*, Upper Saddle River, NJ, USA: Prentice Hall, 2004.
- [50] C.-S. Liu, "Fault detection of rolling element bearings," Seattle, Washington, 2005.
- [51] M. G. Spyridon and I. Chatzisavvas, "Study and construction of an apparatus that automatically monitors vibration and wears in radial ball bearings which are loaded in radial direction," in *International Conference on Signal Processing Systems*, Sydney,

- Australia, 2009.
- [52] M. Satyam, V. Sudhakara Rao and C. Devy, "Cepstrum analysis - an advanced technique in vibration analysis of defects in rotating machinery," *Defense Science Journal*, vol. 44, no. 1, pp. 53-60, 1994.
 - [53] K. Chen, X. Li, F. Wang, T. Wang and C. Wu, "Bearing fault diagnosis using wavelet analysis," in *International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, Chengdu, China, 2012.
 - [54] S. Mallat, *A wavelet tour of signal processing*, 3 ed., Academic Press, 2008.
 - [55] B. Ling and M. Khonsari, "Online tribology ball bearing fault detection and identification," in *Sensors and Systems for Space Applications*, Orlando, Florida, USA, 2007.
 - [56] M. Elmaleeh and N. Saad, "Acoustic emission techniques for early detection of bearing faults using LabVIEW," in *5th International Symposium on Mechatronic and its Applications*, Amman, Jordan, 2008.
 - [57] Y. Yunlong and Z. Zhenxiang, "Fault diagnosis of rolling element bearings based on wavelet analysis," in *International Asia Conference on Informatics in Control, Automation, and Robotics*, Wuhan, China, 2010.
 - [58] D. Wang, Q. Miao, X. Fan and H.-Z. Huang, "Rolling element bearing fault detection using an improved combination of Hilbert and Wavelet transforms," *Journal of Mechanical Science and Technology*, vol. 23, pp. 3292-3301, 2009.
 - [59] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini and A. Bellini, "Fault detection of linear bearings in brushless AC linear motors by vibration analysis," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1684-1694, 2011.
 - [60] J. Liu, "An intelligent system for bearing condition monitoring," Waterloo, Ontario, Canada, 2008.
 - [61] M. Xia, F. Kong and F. Hu, "An approach for bearing fault diagnosis based on PCA and multiple classifier fusion," in *IEEE Joint International Information Technology and Artificial Intelligence Conference*, Chongqing, China, 2011.
 - [62] B. Sreejith, A. Verma and A. Srividya, "Fault diagnosis of rolling element bearings using time-domain features and neural networks," in *3rd International Conference on Industrial and Information Systems*, Kharagpur, India, 2008.
 - [63] L. Jiang, Y. Liu, X. Li and A. Chen, "Degradation assessment and fault diagnosis for roller bearing based on AR model and fuzzy cluster analysis," *Shock and Vibration*, vol. 18, no. 1/2, pp. 127-137, 2011.
 - [64] W. Zuo, D. Zhang and K. Wang, "An assembled matrix distance metric for 2DPCA-based image recognition," *Pattern Recognition Letters*, vol. 27, pp. 210-216, 2006.
 - [65] R. Fontugne, T. Hirotsu and K. Fukuda, "An image processing approach to traffic anomaly detection," in *4th Asian Conference on Internet Engineering*, Bangkok, Thailand, 2008.
 - [66] M. N. Hisyam and C. A. Aziz, "Fault detection for rotating machine using time frequency localization method," Kuantan & Pekan, Pahang, Malaysia, 2010.
 - [67] E. Hu, H. Nosato, H. Sakanashi and M. Murakawa, "Anomaly detection for capsule endoscopy images using higher-order Local Auto Correlation features," in *IEEE International Conference on Systems, Man, and Cybernetics*, Seoul, Korea, 2012.
 - [68] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral

- imagery," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 40, no. 6, pp. 1314-1325, 2002.
- [69] A. Widodo, D. Satrijo, T. Prahasto, G.-M. Lim and B.-K. Choi, "Confirmation of thermal images and vibration signals for intelligent machine fault diagnosis," *International Journal of Rotating Machinery*, vol. 2012, p. 847203, 2012.
 - [70] D. Friedlander, I. Chattopadhyah, A. Ray, S. Phoba and N. Jacobson, "Anomaly prediction in mechanical systems using symbolic dynamics," in *American Control Conference*, Denver, CO, USA, 2003.
 - [71] R. Patton, P. Frank and R. Clark, *Issues of fault diagnosis for dynamic systems*, Springer-Verlag, 2000.
 - [72] N. Merhav, D. Guo and S. Shamai, "Statistical physics of signal estimation in Gaussian noise: theory and examples of phase transitions," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1400-1416, 2010.
 - [73] M. Cuellar and P. Binder, "Reducing noise in discretized time series," *Physical Review E*, vol. 64, no. 4, p. 046211, 2001.
 - [74] B. J. Oommen and E. V. de St. Croix, "Graph partitioning using learning automata," *IEEE Transactions on Computers*, vol. 45, no. 2, pp. 195-208, 1996.
 - [75] T. Chau, "Marginal maximum entropy partitioning yields asymptotically consistent probability density functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 4, pp. 414-417, 2001.
 - [76] M. Kennel and M. Buhl, "Estimating good discrete partitions from observed data: symbolic false nearest neighbors," *Physical Review Letters*, vol. 91, p. 084102, 2003.
 - [77] V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Processing*, vol. 86, no. 11, pp. 3309-3320, 2006.
 - [78] S. Gupta and A. Ray, "Symbolic dynamic filtering for data-driven pattern recognition," in *Pattern Recognition: Theory and Application*, E. A. Zoeller, Ed., Nova Science Publishers, 2007, pp. 17-71.
 - [79] S. Gupta, A. Khatekhate, A. Ray and E. Keller, "Identification of statistical patterns in complex systems via symbolic times series analysis," *ISA Transactions*, vol. 45, no. 4, pp. 477-490, 2006.
 - [80] A. Khatekhate, S. Gupta, A. Ray and R. Patankar, "Anomaly detection in flexible mechanical couplings via symbolic time series analysis," *Journal of Sound and Vibration*, vol. 311, no. 3-5, pp. 608-622, 2008.
 - [81] S. Gupta, A. Ray and E. Keller, "Symbolic time series analysis of ultrasonic signals for fatigue damage monitoring in polycrystalline alloys," *Measurement Science and Technology*, vol. 17, no. 7, pp. 1963-1973, 2006.
 - [82] S. Chakraborty, S. Sarkar, A. Ray and S. Phoba, "Symbolic identification for anomaly detection in aircraft gas turbine engines," in *American Control Conference*, Baltimore, Maryland, USA, 2010.
 - [83] J. Chen, Y.-S. Moon, M.-F. Wong and G. Su, "Palmprint authentication using a symbolic representation of images," *Image and Vision Computing*, vol. 28, no. 3, pp. 343-351, 2010.
 - [84] A. Subbu, A. Srivastav, A. Ray and E. Keller, "Symbolic dynamic filtering for image analysis: theory and experimental validation," *Signal, Image, and Video Processing*, vol. 4,

- no. 3, pp. 319-329, 2010.
- [85] S. C. Chin, A. Ray and V. Rajagopalan, "Symbolic time series analysis for anomaly detection: A comparative evaluation," *Signal Processing*, vol. 85, no. 9, pp. 1859-1868, 2005.
 - [86] J. Lin, E. Keogh, S. Lonardi and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego, CA, USA, 2003.
 - [87] J. Li, A. Najmi and R. Gray, "Image classification by a two-dimensional hidden Markov model," *Transactions on Signal Processing*, vol. 48, no. 2, pp. 517-533, 2000.
 - [88] R. Yan and R. X. Gao, "Multi-scale enveloping spectrogram for vibration analysis in bearing defect diagnosis," *Tribology International*, vol. 42, no. 2, pp. 293-302, 2008.
 - [89] R. Klein, E. Masad, E. Rudyk and I. Winkler, "Bearing diagnostics using image processing methods," Gilon, Isreal, 2011.
 - [90] E. Keogh, J. Lin, A. Fu and W.-C. Wai-Chee, "HOT SAX: efficiently finding the the most unusual time series subsequence," in *5th IEEE International Conference on Data Mining*, Houston, Texas, USA, 2005.
 - [91] S. Kasetty, C. Stafford, G. P. Walker, X. Wang and E. Keogh, "Real-time classification of streaming sensor data," in *20th IEEE International Conference on Tools with Artificial Intelligence*, Dayton, OH, USA, 2008.
 - [92] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: a survey and empirical demonstration," *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 349-371, 2003.
 - [93] N. Swansson and S. Favaloro, "Applications of vibration analysis to the condition monitoring of rolling element bearings," Melbourne, Victoria, Australia, 1984.
 - [94] Y.-C. Lai, E. Boltt and C. Grebogi, "Communicating with chaos using two-dimensional symbolic dynamics," *Physics Letters A*, vol. 255, pp. 75-81, 1999.
 - [95] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115-1130, 2004.
 - [96] Y. Liu, D.-p. Bao and Z.-h. Yang, "Application research of a new symbolic approximation method - SAX in time series mining," *Computer Engineering and Applications*, vol. 42, no. 27, 2006.
 - [97] M. Zoumboulakis and G. Roussos, "Escalation: complex event detection in wireless sensor networks," in *2nd European Conference on Smart Sensing and Context*, Kendal, England, 2007.
 - [98] J. Norris, *Markov Chains*, Cambridge University Press, 1998.
 - [99] R. Duda, P. Hart and D. Stork, *Pattern classification*, 2 ed., John Wiley & Sons, 2001.
 - [100] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, Oxford University Press, 2001.
 - [101] A. Kadrolkar and R. Gao, "Statistical analysis of symbol sequence distributions for machine condition monitoring," in *34th Annual Conference of IEEE Industrial Electronics.*, Orlando, FL, USA, 2008.
 - [102] L. Wei, N. Kumar, V. Lolla, E. Keogh, S. Lonardi and C. A. Ratananmahatana,

- "Assumption-free anomaly detection in time-series," in *17th International Conference on Scientific and Statistical Database Management*, Santa Barbara, CA, USA, 2005.
- [103] B. Kulahcioglu, S. Ozdemir and B. Kumova, "Application of symbolic Piecewise Aggregate Approximation (PAA) analysis to ECG signals," in *17th IASTED International Conference on Applied Simulation and Modelling*, Corfu, Greece, 2008.
- [104] S. Sarkar, K. Mukherjee, X. Jin, D. Singh and A. Ray, "Optimization of symbolic feature extraction for pattern classification," *Signal Processing*, vol. 92, no. 3, pp. 625-635, 2012.
- [105] V. Bevilacqua, L. Cariello, C. Gaetano, D. Daleno and G. Mastronardi, "Pseudo 2D Hidden Markov Models for face recognition using neural network coefficients," *Soft Computing*, vol. 12, pp. 615-621, 2008.
- [106] D. Van Hieu and S. Nitsuwat, "Image processing and trajectory feature extraction based on Hidden Markov Models for sign language recognition," in *9th International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, Bangkok, Thailand, 2008.
- [107] P. Tarazaga, M. Raydan and A. Hurman, "Perron–Frobenius theorem for matrices with some negative entries," *Linear Algebra and its Applications*, vol. 328, no. 1-3, pp. 57-68, 2001.
- [108] P. Kontkanen, J. Lahtinen, P. Myllymaki and H. Tirri, "An unsupervised Bayesian distance metric," *Advances in Case-Based Reasoning Lecture Notes in Computer Science*, vol. 1898, pp. 148-160, 2000.
- [109] V. Cappellini, H. Sommers, W. Bruzda and K. Zyczkowski, "Random bistochastic matrices," *Journal of Physics A: Mathematical and Theoretical*, vol. 42, p. 365209, 2009.
- [110] C. Li and G. Yu, "A new statistical model for rolling element bearing fault signals based on alpha-stable distribution," in *2nd International Conference on Computer Modeling and Simulation*, Sanya, China, 2010.
- [111] J. S. Mitchell, "From vibration measurements to condition based monitoring," *Sound and Vibration, 40th Anniversary Issue*, pp. 62-75, 2007.
- [112] W. Reimche, U. Sudmerson, O. Pietsch, C. Scheer and F.-W. Bach, "Basics of vibration monitoring for fault detection and process control," Rio de Janeiro, Brazil, 2003.
- [113] A. Sant'Anna and N. Wickstrom, "Symbolization of time-series: an evaluation of SAX, Persist, and ACA," in *4th International Congress on Image and Signal Processing*, Shanghai, China, 2011.
- [114] S. Sassi, B. Badri and M. Thomas, "A numerical model to predict damaged bearing vibrations," *Journal of Vibration and Control*, vol. 13, no. 11, pp. 1603-1628, 2007.
- [115] H. Fang and B. Yang, "Modeling, synthesis, and dynamic analysis of complex flexible rotor systems," *Journal of Sound and Vibration*, vol. 211, no. 4, pp. 571-592, 1998.
- [116] C. Zhang, J. Qui, S. Danyluk, T. Kurfess and S. Liang, "Impact dynamics modeling of bearing vibration for defect size estimation," *International Journal of Condition Monitoring and Diagnostic Engineering Management*, vol. 3, no. 3, pp. 37-42, 2000.
- [117] N. Nikolaou and I. Antoniadis, "Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted morlet wavelets," *Mechanical Systems and Signal Processing*, vol. 16, no. 4, pp. 677-694, 2002.
- [118] B. Hamrock, *Fundamentals of fluid film lubrication*, McGraw-Hill, 1994.

- [119] F. Gioia and C. N. Lauro, "Principal component analysis of interval data," *Computational Statistics*, vol. 21, pp. 343-363, 2006.
- [120] S. Baronti, R. Carla, S. Sigismondi and L. Alparone, "Principal component analysis for change detection on polarimetric multitemporal SAR data," in *IEEE International Geoscience and Remote Sensing Symposium*, Singapore, 1994.
- [121] A. Khatekhate, A. Ray, E. Keller, S. Gupta and S. Chin, "Symbolic time-series analysis for anomaly detection in mechanical systems," *Transactions on Mechatronics*, vol. 11, no. 4, pp. 439-447, 2006.
- [122] L. Garibaldi, A. Fasana and S. Marchesiello, "An experimental rig for damage detection of rolling bearings," *Key Engineering Materials*, vol. 293/294, pp. 509-516, 2005.

A ISO 10816 VIBRATION STANDARDS

VIBRATION SEVERITY PER ISO 10816					
Machine		Class I small machines	Class II medium machines	Class III large rigid foundation	Class IV large soft foundation
	in/s mm/s				
Vibration Velocity Vrms	0.01 0.28				
	0.02 0.45				
	0.03 0.71		good		
	0.04 1.12				
	0.07 1.80				
	0.11 2.80		satisfactory		
	0.18 4.50				
	0.28 7.10		unsatisfactory		
	0.44 11.2				
	0.70 18.0				
	0.71 28.0		unacceptable		
	1.10 45.0				

Figure A.1. ISO 10816 vibration severity chart for machine vibration.

B CWRU EXPERIMENTAL TEST CONDITIONS

Table B.1. Normal test parameters and conditions.

Experiment #	Motor load (hp)	Motor speed (rpm)
1	0	1797
2	1	1772
3	2	1750
4	3	1730

Table B.2. Faulty test parameters

Experiment #	Motor load (hp)	Motor speed (rpm)	Fault location	Fault size (in)
5	0	1797	Outer race	0.007
6	1	1772	Outer race	0.007
7	2	1750	Outer race	0.007
8	3	1730	Outer race	0.007
9	0	1797	Outer race	0.014
10	1	1772	Outer race	0.014
11	2	1750	Outer race	0.014
12	3	1730	Outer race	0.014
13	0	1797	Outer race	0.021
14	1	1772	Outer race	0.021
15	2	1750	Outer race	0.021
16	3	1730	Outer race	0.021
17	0	1797	Inner race	0.007
18	1	1772	Inner race	0.007
19	2	1750	Inner race	0.007
20	3	1730	Inner race	0.007
21	0	1797	Inner race	0.014
22	1	1772	Inner race	0.014
23	2	1750	Inner race	0.014
24	3	1730	Inner race	0.014
25	0	1797	Inner race	0.021
26	1	1772	Inner race	0.021
27	2	1750	Inner race	0.021
28	3	1730	Inner race	0.021
29	0	1797	Ball	0.007
30	1	1772	Ball	0.007
31	2	1750	Ball	0.007
32	3	1730	Ball	0.007
33	0	1797	Ball	0.014
34	1	1772	Ball	0.014
35	2	1750	Ball	0.014
36	3	1730	Ball	0.014
37	0	1797	Ball	0.021
38	1	1772	Ball	0.021
39	2	1750	Ball	0.021
40	3	1730	Ball	0.021

c NUMERIC MDOF SIMULINK MODEL

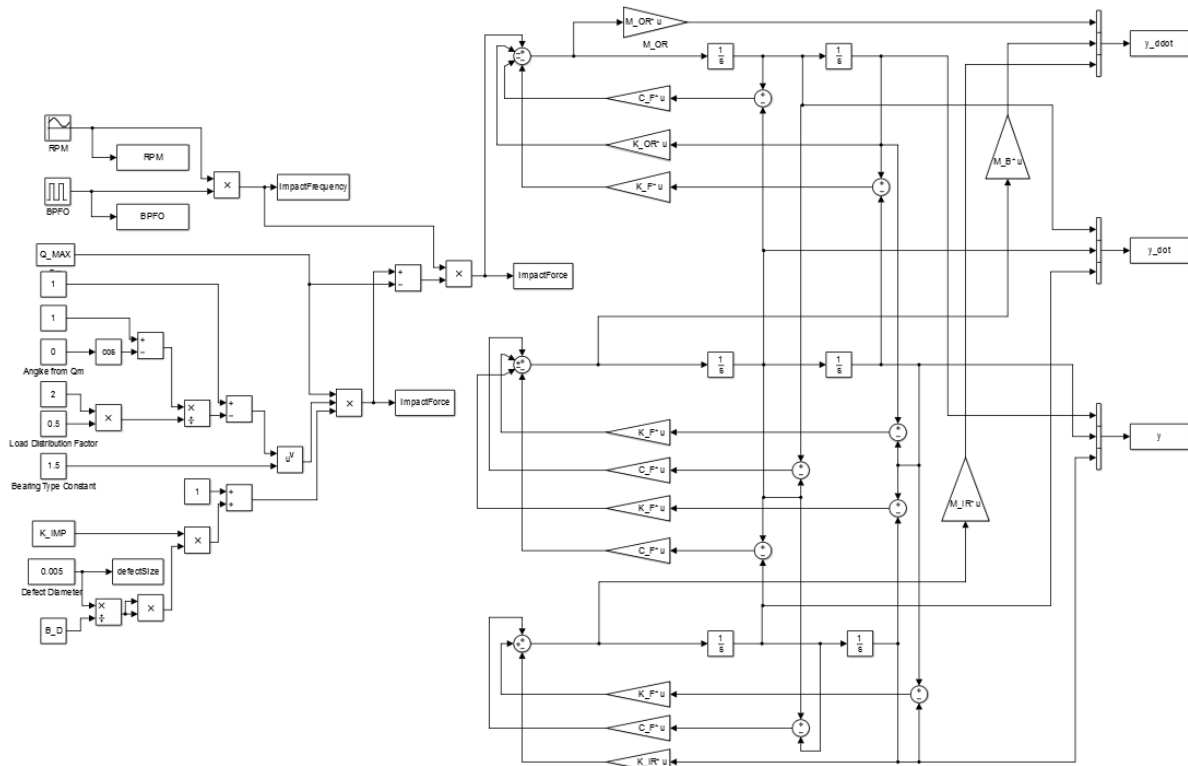


Figure C.1. (a) SIMULINK model for developing propagating OR vibration signature.

D CALCULATION OF STIFFNESS AND DAMPING MATRIX

This model assumes the fluid film obeys Newtonian viscous effects, has negligible inertial effects, is incompressible, and an unwavering viscosity and pressure through time.

From the SKF 6205 2RS-JEM bearing, the parameters and geometries are:

$$\begin{aligned}
 OD &= 0.052 \text{ m} \\
 ID &= 0.025 \text{ m} \\
 L &= 0.015 \text{ m} \\
 2P_r/\varphi_B &= 4.9 \\
 P_d = 2P_r &= (OD + ID)/2 = 0.0385 \text{ m} \\
 \varphi_B = P_d/4.9 &= 0.00787 \text{ m} \\
 \alpha &= 0 \text{ rad} \\
 J_b &= 9 \text{ balls} \\
 \psi_m &= 0 \text{ rad}
 \end{aligned}$$

Under the operating conditions described in Section 3.5:

$$\begin{aligned}
 Q_m &= 450 \text{ N} \\
 \varepsilon &= 0.5 \\
 w_{IR} &= 1750 \text{ rpm} = 183.26 \text{ rad/s} \\
 w_{OR} &= 0 \text{ rpm} = 0 \text{ rad/s}
 \end{aligned}$$

From previously FE validated stiffness values:

$$\begin{aligned}
 K_{IR} &= 2.48(10)^8 \text{ N/m} \\
 K_{OR} &= 5.96(10)^7 \text{ N/m} \\
 K_B &= \infty \text{ N/m}
 \end{aligned}$$

Calculation of equivalent mass values:

$$\begin{aligned}
 w_{n_{IR}} &= 10593 \text{ Hz} \\
 w_{n_{OR}} &= 4102 \text{ Hz} \\
 M_{IR} &= K_{IR}/w_{n_{IR}}^2 = 2.21 \text{ kg} \\
 M_{OR} &= K_{OR}/w_{n_{OR}}^2 = 3.54 \text{ kg} \\
 M_B &= \rho_{STEEL} V_B = (7800 \text{ kg/m}^3)(2.09\varphi_B)^3 = 0.002 \text{ kg}
 \end{aligned}$$

Next, the damping values are calculated from EHD theory. c is diametric clearance where c_o and c_m are found from the manufacturer. Due to the relative movement of the rolling elements, it is suggested to use the inner race frequency for the ‘journal’ spin frequency w_j . P_B is the average bearing pressure in the projected area of the journal. The Sommerfield number S_F is the characteristic bearing number [22], assuming a $\mu_F = 4(10)^{-6}$ reyn absolute viscosity at 150°F.

$$\begin{aligned}
c_0 &= 0.000013 \text{ m} \\
c_m &= 0.000028 \text{ m} \\
c &= (c_0 + c_m)/2 = 0.0000205 \text{ m} \\
w_j &= w_{IR} = 29.17 \text{ r/s} \\
P &= Q_m/(ID/2)L = 2.4(10)^6 \text{ N/m}^2 \\
S_F &= ((ID/2)/c)^2(\mu_F w_j/P) = 0.1348
\end{aligned}$$

Knowing a $L:ID$ ratio of 0.65, the minimum bearing eccentricity becomes:

$$\begin{aligned}
h_0/c &= 0.34 \\
\varepsilon_0 &= 1 - (h_0/c) = 0.66
\end{aligned}$$

The dimensionless stiffness and damping coefficients of the fluid film are found from EHD theory:

$$\begin{aligned}
\varphi_0 &= \tan^{-1}(\pi(1 - \varepsilon_0^2)^{1/2})/4\varepsilon_0 = 0.8084 \text{ rad} \\
\emptyset &= \frac{(1 - \varepsilon_0^2)^2}{\varepsilon_0(16\varepsilon_0^2 + \pi^2(1 - \varepsilon_0^2))^{1/2}} \cos^2 \varphi_0 = 0.0758
\end{aligned}$$

$$K = \emptyset \left(\frac{\varepsilon_0}{(1 - \varepsilon_0^2)^2} \sin^2 \varphi_0 + \frac{3\pi\varepsilon_0^2}{4(1 - \varepsilon_0^2)^{5/2}} \sin\varphi_0 \cos\varphi_0 + \frac{2\varepsilon_0(1 + \varepsilon_0^2)}{(1 - \varepsilon_0^2)^3} \cos^2 \varphi_0 \right) = 0.6749$$

$$C = \emptyset \left(\frac{\pi}{2(1 - \varepsilon_0^2)^{3/2}} \sin^2 \varphi_0 + \frac{4\pi\varepsilon_0}{(1 - \varepsilon_0^2)^2} \sin\varphi_0 \cos\varphi_0 + \frac{\pi(1 + 2\varepsilon_0^2)}{2(1 - \varepsilon_0^2)^{5/2}} \cos^2 \varphi_0 \right) = 0.9537$$

$$\begin{aligned}
K_F &= KQ_m/h_0 = 11,193.6 \text{ N/m} \\
C_F &= CQ_m w_{IR}/h_0 = 542.3 \text{ N.s/m}
\end{aligned}$$

E 1D AND 2D SAX MATLAB CODES

```
function [return] = sax(X,B,WINDOW,NOVERLAP)
%SAXIT returns a Symbolic Aggregate Approximation of a zero-averaged array
%
%      X      = input array                \\ Required Value
%      B      = resolution, # of breakpoints, 3-20 \\ Required value
%      WINDOW  = cardinality, # of data-points to average
%      NOVERLAP = overlap value, # of data-points to overlap

%% Input processing and defaulting
if ~exist('X','var') || ~exist('B','var'), error('No array or resolution inputted'); end

if ~exist('WINDOW','var') || isempty(WINDOW), WINDOW = 1;      end
if ~exist('NOVERLAP','var') || isempty(NOVERLAP), NOVERLAP = 0; end
WINDOW = floor(WINDOW);
NOVERLAP = floor(NOVERLAP);
B = floor(B);
if WINDOW < 1 || WINDOW > floor(length(X)/2), error('\nWINDOW must be a positive integer 1 <= WINDOW <= N/2 (N = array length)'); end
if NOVERLAP < 0 || NOVERLAP > floor(length(X)/2), error('\nNOVERLAP must be a positive integer 0 <= WINDOW <= N/2 (N = array length)'); end
if NOVERLAP >= WINDOW, error('\nNumber of overlap points exceeds window size'); end
if B < 3 || B > 20, error('\nUser-specified value of %i breakpoints is invalid:\nNumber of breakpoints must be between 3 and 20\nSee [help saxit] for more information',floor(B)); end
R = length(size(X));    %# of dimensions
if R > 2, error('\nArray must be 1-D or 2-D, you specified a %i-dimensional array.',R); end

%% Setup Z-score table
bpArray = ...
[-0.43 -0.67 -0.84 -0.97 -1.07 -1.15 -1.22 -1.28 -1.34 -1.38 -1.43 -1.47 -1.5 -1.53 -1.56 -1.59 -1.62 -1.64;
 0.43 0 -0.25 -0.43 -0.57 -0.67 -0.76 -0.84 -0.91 -0.97 -1.02 -1.07 -1.11 1.15 -1.19 -1.22 -1.25 -1.28;
 0 0.67 0.25 0 -0.18 -0.32 -0.43 -0.52 -0.6 -0.67 -0.74 -0.79 -0.84 -0.89 0.93 -0.97 -1 -1.04;
 0 0 0.84 0.43 0.18 0 -0.14 -0.25 -0.35 -0.43 -0.5 -0.57 -0.62 -0.67 -0.72 0.76 -0.8 -0.84;
 0 0 0 0.97 0.57 0.32 0.14 0 -0.11 -0.21 -0.29 -0.37 -0.43 -0.49 -0.54 -0.59 -0.63 -0.67;
 0 0 0 1.07 0.67 0.43 0.25 0.11 0 -0.1 -0.18 -0.25 -0.32 -0.38 -0.43 -0.48 -0.52;
 0 0 0 0 0 1.15 0.76 0.52 0.35 0.21 0.1 0 -0.08 -0.16 -0.22 -0.28 -0.34 -0.39;
 0 0 0 0 0 1.22 0.84 0.6 0.43 0.29 0.18 0.08 0 -0.07 -0.14 -0.2 -0.25;
 0 0 0 0 0 1.28 0.91 0.67 0.5 0.37 0.25 0.16 0.07 0 -0.07 -0.13;
 0 0 0 0 0 0 1.34 0.97 0.74 0.57 0.43 0.32 0.22 0.14 0.07 0;
 0 0 0 0 0 0 1.38 1.02 0.79 0.62 0.49 0.38 0.28 0.2 0.13;
 0 0 0 0 0 0 1.43 1.07 0.84 0.67 0.54 0.43 0.34 0.25;
 0 0 0 0 0 0 1.47 1.11 0.89 0.72 0.59 0.48 0.39;
 0 0 0 0 0 0 1.5 1.15 0.93 0.76 0.63 0.52;
 0 0 0 0 0 0 1.53 1.19 0.97 0.8 0.67;
 0 0 0 0 0 0 1.56 1.22 1 0.84;
 0 0 0 0 0 0 1.59 1.25 1.04;
 0 0 0 0 0 0 1.62 1.28;
 0 0 0 0 0 0 1.64];

%% 1-Dimensional Array
if size(X,1) == 1 || size(X,2) == 1
%Window data, does not account for cases where mod(N/n)~=0
myArray = zeros(1,ceil(length(X)/(WINDOW-NOVERLAP)));
if WINDOW ~= 1
    k = 1;
    for j = 1:WINDOW-NOVERLAP:length(X)-WINDOW+1
        myArray(k) = sum(X(j:j+WINDOW-1))/WINDOW;
        k = k+1;
    end
else myArray = X;
end
%Zero-average the processed array
Zscores = (myArray - (sum(myArray)/length(myArray)))/std(myArray);
myLetters = ['a' 'b' 'c' 'd' 'e' 'f' 'g' 'h' 'i' 'j' 'k' 'l' 'm' 'n' 'o' 'p' 'q' 'r' 's' 't'];
saxArray = zeros(length(Zscores),1);
%Final step, assign letters and return
for j=1:length(myArray)
    if Zscores(j) < bpArray(1,B-2), saxArray(j) = myLetters(1); else
        for b=2:B-1
```

```

        if Zscores(j) >= bpArray(b-1,B-2) && Zscores(j) < bpArray(b,B-2),
            saxArray(j) = myLetters(b); break, end
        end
    end
    if Zscores(j) >= bpArray(B-1,B-2), saxArray(j) = myLetters(B); end
end
end

%% 2-Dimensional Array
if size(X,1) ~= 1 && size(X,2) ~= 1
    %Window chunks of data, does not account for cases where mod(N/n)~=0
    myArray = zeros(ceil(size(X,1)/(WINDOW-NOVERLAP)),ceil(size(X,2)/(WINDOW-NOVERLAP)));
    if WINDOW ~= 1
        x = 1;
        y = 1;
        for k = 1:WINDOW-NOVERLAP:size(X,1)-WINDOW+1
            for j = 1:WINDOW-NOVERLAP:size(X,2)-WINDOW+1
                myArray(x) = sum(sum(X(k:k+WINDOW-1,j:j+WINDOW-1)))/WINDOW^2;
                x = x+1;
            end
            y = y+1;
        end
    else myArray = X;
    end

    %Zero-average the processed array
    Zscores = (myArray - (sum(sum(myArray)) / length(myArray(:)))) / std(myArray(:));
    myLetters = ['a' 'b' 'c' 'd' 'e' 'f' 'g' 'h' 'i' 'j' 'k' 'l' 'm' 'n' 'o' 'p' 'q' 'r' 's' 't'];
    saxArray = zeros(size(Zscores));
    %Final step, assign letters and return
    for k = 1:1:size(myArray,1)
        for j = 1:1:size(myArray,2)
            if Zscores(k,j) < bpArray(1,B-2), saxArray(k,j) = myLetters(1); else
                for b = 2:1:B-1
                    if Zscores(k,j) >= bpArray(b-1,B-2) && Zscores(k,j) < bpArray(b,B-2),
                        saxArray(k,j) = myLetters(b); break, end
                    end
                end
                if Zscores(k,j) >= bpArray(B-1,B-2), saxArray(k,j) = myLetters(B); end
            end
        end
    end
end
return = saxArray;

```

F 2D HMM GENERATION MATLAB CODES

```

[~,~,~,P]      = spectrogram(SIGNAL,SPECWINDOW,SPECOVERLAP);
specImage      = (10*log10(P));
symSpec        = saxit(specImage,B,WINDOW,NOVERLAP);
if SIZEIGHBORHOOD == 1
    stateImage   = symSpec;
    dataFreq     = getBayesTable(stateImage,B);
elseif SIZEIGHBORHOOD == 3
    imageHeight  = length(symSpec(:,1));
    allStates    = getStates(SIZEIGHBORHOOD,B);
    pixels       = zeros(imageHeight-2,(ANOMALYWINDOW / SPECWINDOW) - 2,
        SIZEIGHBORHOOD^2);
    states       = zeros(imageHeight-2,(ANOMALYWINDOW/SPECWINDOW)2,B);
    stateImage    = zeros(imageHeight-2,(ANOMALYWINDOW/SPECWINDOW)-2);
    for r = 2:imageHeight-1
        for s = 2:(ANOMALYWINDOW/SPECWINDOW)-1
            pixels (r-1,s-1,1) = symSpec (r-1,s-1);
            pixels (r-1,s-1,2) = symSpec (r-1,s);
            pixels (r-1,s-1,3) = symSpec (r-1,s+1);
            pixels (r-1,s-1,4) = symSpec (r,s-1);
            pixels (r-1,s-1,5) = symSpec (r,s);
            pixels (r-1,s-1,6) = symSpec (r,s+1);
            pixels (r-1,s-1,7) = symSpec (r+1,s-1);
            pixels (r-1,s-1,8) = symSpec (r+1,s);
            pixels (r-1,s-1,9) = symSpec (r+1,s+1);
            states (r-1,s-1,1) = sum((pixels (r-1,s-1,:)) == 1);
            states (r-1,s-1,2) = sum((pixels (r-1,s-1,:)) == 2);
            states (r-1,s-1,3) = sum((pixels (r-1,s-1,:)) == 3);
            for d = 1:length(allStates(1,:))
                if (states (r-1,s-1,1) - allStates(1,d)) == 0 && ...
                    (states (r-1,s-1,2) - allStates(2,d)) == 0 && ...
                    (states (r-1,s-1,3) - allStates(3,d)) == 0
                    stateImage (r-1,s-1) = d; break
                end
            end
        end
    end
    dataFreq = getBayesTable(stateImage,length(allStates(1,:)));
end
[W,~]        = eig(dataFreq. '); W = conj(W);
data         = abs(W(:,1));

```

```

function [allStates] = getStates(sizeNeighborhood,B)

{
if B == 3
    if sizeNeighborhood == 1,      allCombs = allcomb([1 2 3]);
    elseif sizeNeighborhood == 3, allCombs = allcomb([1 2 3],[1 2 3],[1 2 3],[1 2 3],[1 2 3],[1 2
        3],[1 2 3],[1 2 3],[1 2 3]); end
elseif B == 2
    if sizeNeighborhood == 1,      allCombs = allcomb([1 2]);
    elseif sizeNeighborhood == 3,
        allCombs = allcomb([1 2],[1 2],[1 2],[1 2],[1 2],[1 2],[1 2],[1 2],[1 2],[1 2]);
    end
end

States = zeros(B,length(allCombs(:,1)));
for k = 1:length(allCombs(:,1))
    States(1,k) = sum((allCombs(k,:) == 1));
    States(2,k) = sum((allCombs(k,:) == 2));
    States(3,k) = sum((allCombs(k,:) == 3));
end

finalStates = zeros(3,length(States(1,:)));
for i = 1:length(States(1,:))-1
    for j = i+1:length(States(1,:))
        if (States(1,j) - States(1,i)) == 0 && ...
            (States(2,j) - States(2,i)) == 0 && ...
            (States(3,j) - States(3,i)) == 0
            finalStates(:,i) = 0; break
        else
            finalStates(:,i) = States(:,i);
            finalStates(:,j) = States(:,j);
        end
    end
end
allStates = finalStates(:,any(finalStates));
end
}

function [bayesTable] = getBayesTable(myData,B)
%% 1D Vector
if size(myData,1)==1 || size(myData,2) == 1, bayesTable = zeros(N,N);
    for k = 1: numel(myData)-1
        bayesTable(myData(k),myData(k+1)) = bayesTable(myData(k),myData(k+1)) + 1;
    end
    bayesTable = bayesTable + 0.001;
    for b = 1:1:N
        bayesTable(b,:) = bayesTable(b,:) / sum(bayesTable(b,:));
    end
end

%% 2D Image
if size(myData,1)~=1 && size(myData,2)~=1, bayesTable = zeros(N,N);
    O = length(myData(:,1)); P = length(myData(1,:));
    for o = 2:O-1
        for p = 2:P-1
            bayesTable(myData(o,p),myData(o+1,p)) = bayesTable(myData(o,p),myData(o+1,p)) + 1;
            bayesTable(myData(o,p),myData(o,p+1)) = bayesTable(myData(o,p),myData(o,p+1)) + 1;
            bayesTable(myData(o,p),myData(o+1,p+1)) = bayesTable(myData(o,p),myData(o+1,p+1)) + 1;
        end
    end
    bayesTable = bayesTable + 0.001;
    for b = 1:N
        bayesTable(b,:) = bayesTable(b,:) / sum(bayesTable(b,:));
    end
end

```

G EXPERIMENTAL DATA ANOVA TABLES

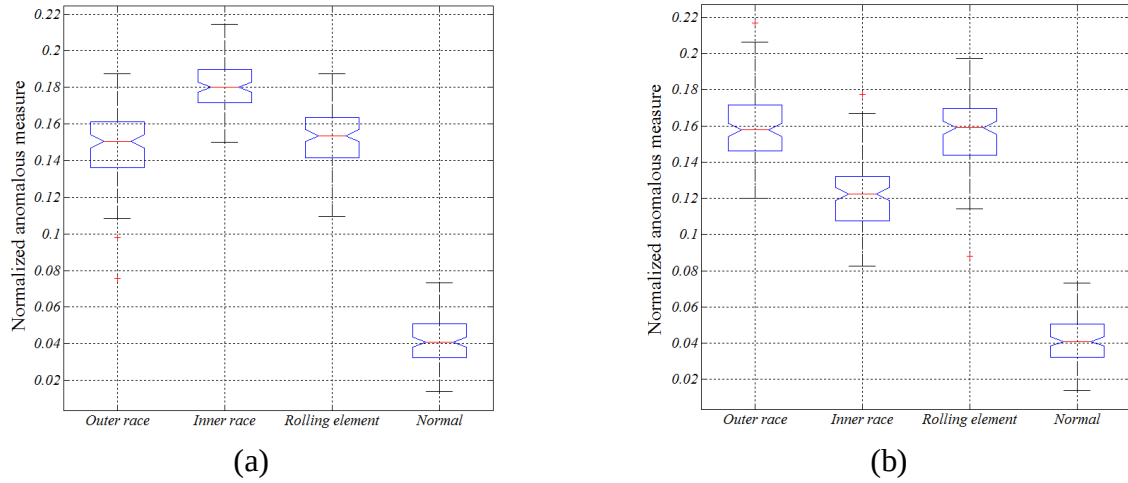


Figure G.1. 2D anomalous measure distributions of (a) 0.007 in and (b) 0.021 in faults.

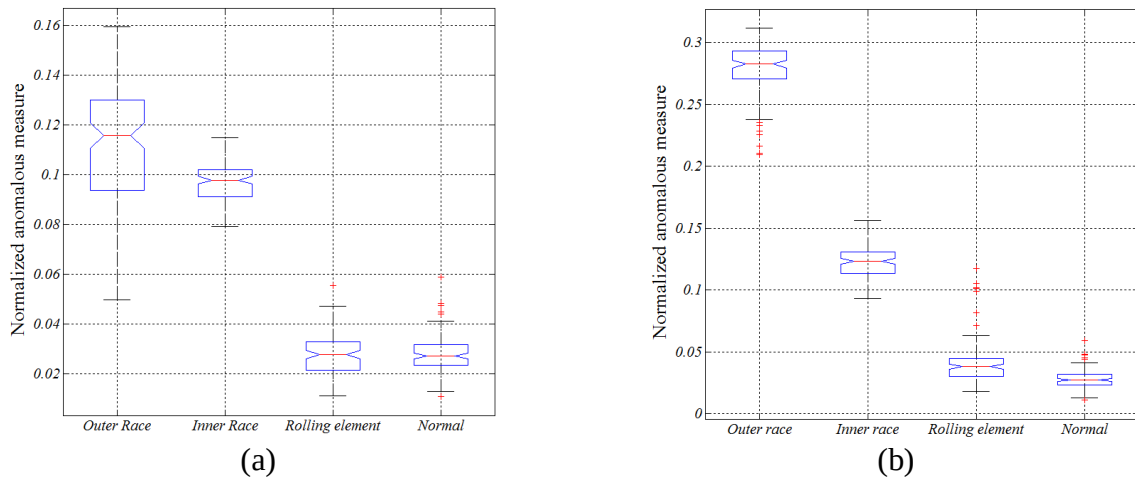


Figure G.2. 1D anomalous measure distributions of (a) 0.007 in and (b) 0.021 in faults.

Table G.1. ANOVA table from all data in Figure G.1 (a).

Source	SS	df	MS	F	Prob > F
Columns	1.34009	3	0.4467	1996.99	3.53479(10) ⁻²⁶⁹
Error	0.10647	476	0.00022		
Total	1.44656	479			

Table G.2. ANOVA table from all data in Figure G.1 (b).

Source	SS	df	MS	F	Prob > F
Columns	1.09995	3	0.36665	1164.19	1.02263(10) ⁻²¹⁸
Error	0.14991	476	0.00031		
Total	1.24986	479			

Table G.3. ANOVA table comparing the three fault locations in Figure G.1 (b).

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Prob > F</i>
Columns	0.10946	2	0.05473	149.25	7.82575(10) ⁻⁴⁸
Error	0.13092	357	0.00037		
Total	0.24038	359			

Table G.4. ANOVA table comparing the OR and IR faults in Figure G.1 (b).

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Prob > F</i>
Columns	0.09063	1	0.09063	249.17	6.84948(10) ⁻⁴⁹
Error	0.08656	238	0.00036		
Total	0.17719	239			

Table G.5. ANOVA table comparing the OR and RE fault in Figure G.1 (b).

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Prob > F</i>
Columns	0.001	1	0.001	2.57	0.1105
Error	0.09299	238	0.00039		
Total	0.09399	239			

Table G.6. ANOVA table comparing the RE and IR fault in Figure G.1 (b).

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Prob > F</i>
Columns	0.07256	1	0.07256	209.89	1.58199 (10) ⁻³⁴
Error	0.08228	238	0.00035		
Total	0.15485	239			

H BEARING RIG DEVELOPMENT

The current experimental bearing test bed contains a one T-slot aluminum frame base 4 feet in length based from the design of [122]. This supports the 1.5 hp (1750 rpm max) LESSON electric motor, and two aluminum bearing housings. The output of this motor is coupled to a ½” shaft with a flexible coupler. The shaft is held in place by the two bearing housings, thus allowing motor vibration to be incorporated into the model. The two bearing housings allow for the bearing’s inner ring to rotate under a fixed outer ring. The specific bearings can include any manufacturer type, given a 2 in OD. Two piezoelectric accelerometers are also placed on the top of the bearing housings to transmit the bearing’s acceleration signal. An optical sensor is also used to gauge shaft speed during experiments. The sensor detects a thin film of reflective tape on the rotating shaft at successive revolutions. A LabVIEW program and National Instruments’ DAQ-PAD recorded preliminary data from the accelerometers and the optical speed sensor to verify experimental setup.

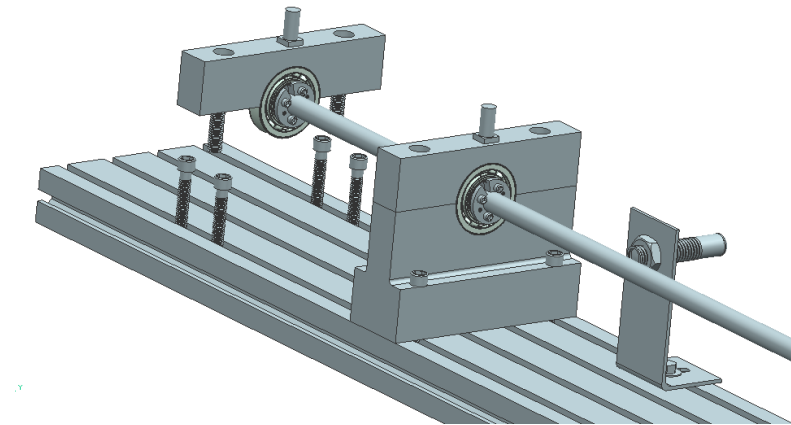


Figure H.1. CAD model of current experimental bearing rig.

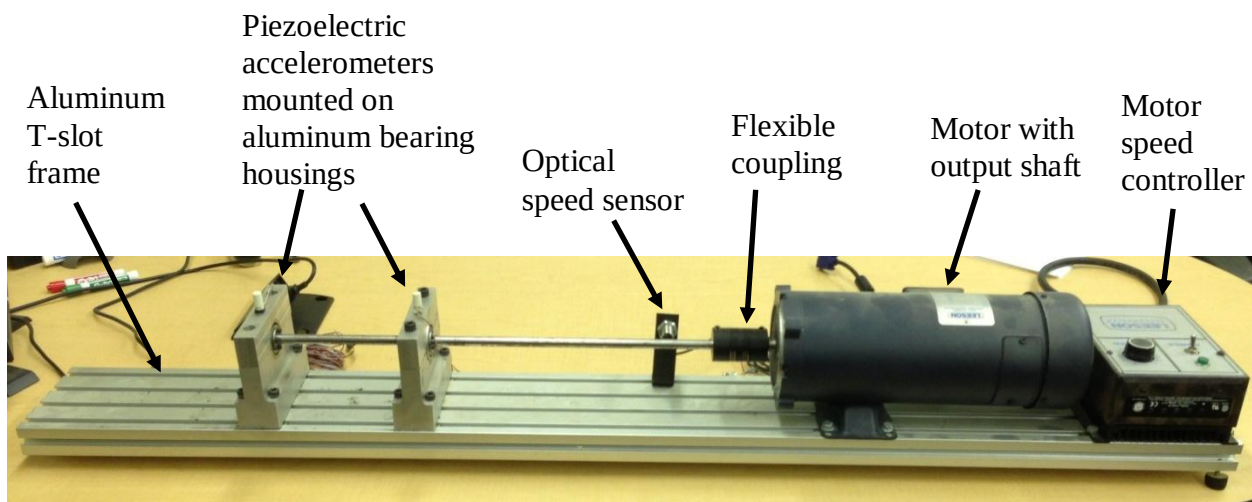


Figure H.2. Current experimental bearing test bed.