A DIGITAL TRIPLET FOR UTILIZING OFFLINE ENVIRONMENTS TO TRAIN CONDITION MONITORING SYSTEMS FOR ROLLING ELEMENT BEARINGS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mechanical Engineering

by
Ethan SangWon Koh Wescoat
December 2023

Accepted by:
Dr. Laine Mears, Committee Chair
Dr. Adam Hoover
Dr. Cameron Turner
Dr. John Wagner

Abstract

Manufacturing competitiveness is related to making a quality product while incurring the lowest costs. Unexpected downtime caused by equipment failure negatively impacts manufacturing competitiveness due to the ensuing defects and delays caused by the downtime. Manufacturers have adopted condition monitoring (CM) techniques to reduce unexpected downtime to augment maintenance strategies. The CM adoption has transitioned maintenance from Breakdown Maintenance (BM) to Condition-Based Maintenance (CbM) to anticipate impending failures and provide maintenance actions before equipment failure. CbM is the umbrella term for maintenance strategies that use condition monitoring techniques such as Preventive Maintenance (PM) and Predictive Maintenance (PdM). Preventive Maintenance involves providing periodic checks based on either time or sensory input. Predictive Maintenance utilizes continuous or periodic sensory inputs to determine the machine health state to predict the equipment failure.

The overall goal of the work is to improve bearing diagnostic and prognostic predictions for equipment health by utilizing surrogate systems to generate failure data that represents production equipment failure, thereby providing training data for condition monitoring solutions without waiting for real world failure data. This research seeks to address the challenges of obtaining failure data for CM systems by incorporating a third system into monitoring strategies to create a Digital Triplet (DTr) for condition monitoring to increase the amount of possible data for condition monitoring. Bearings are a critical component in rotational manufacturing systems with wide application to other industries outside of manufacturing, such as energy and defense. The reinvented DTr system considers three components: the physical, surrogate, and digital systems. The physical system represents the real-world application in production that cannot fail. The surrogate system represents a physical component in a test system in an offline environment where data is generated to fill in gaps from data unavailable in the real-world system. The digital system is the CM system, which

provides maintenance recommendations based on the ingested data from the real world and surrogate systems.

In pursuing the research goal, a comprehensive bearing dataset detailing these four failure modes over different collection operating parameters was created. Subsequently, the collections occurred under different operating conditions, such as speed-varying, load-varying, and steady-state. Different frequency and time measures were used to analyze and identify differentiating criteria between the different failure classes over the differing operating conditions. These empirical observations were recreated using simulations to filter out potential outliers. The outputs of the physical model were combined with knowledge from the empirical observations to create "spectral deltas" to augment existing bearing data and create new failure data that resemble similar frequency criteria to the original data. The primary verification occurred on a laboratory-bearing test stand. A conjecture is provided on how to scale to a larger system by analyzing a larger system from a local manufacturer.

From the subsequent analysis of machine learning diagnosis and prognosis models, the original and augmented bearing data can complement each other during model training. The subsequent data substitution verifies that bearing data collected under different operating conditions and sizes can be substituted between different systems. Ostensibly, the full formulation of the digital triplet system is that bearing data generated at a smaller size can be scaled to train predictive failure models for larger bearing sizes. Future work should consider implementing this method for other systems outside of bearings, such as gears, non-rotational equipment, such as pumps, or even larger complex systems, such as computer numerically controlled machine tools or car engines. In addition, the method and process should not be restricted to only mechanical systems and could be applied to electrical systems, such as batteries. Furthermore, an investigation should consider further data-driven approximations to specific bearing characteristics related to the stiffness and damping parameters needed in modeling. A final consideration is for further investigation into the scalability quantities within the data and how to track these changes through different system levels.

Dedication

For Noah, HeeJung, and Andy, who never failed to lift me up from day one when the journey was difficult.

Acknowledgments

Table of Contents

List of Tables

List of Figures

xviii

xix

Nomenclature

CM	Condition Monitoring
BM	Breakdown Maintenance
CBM	Condition-Based Maintenance
PM	Preventive Maintenance
PdM	Predictive Maintenance
PFailM	Purposeful Failure Methodology
1IR	First Industrial Revolution
2IR	Second Industrial Revolution
3IR	Third Industrial Revolution
4IR	Fourth Industrial Revolution
5IR	Fifth Industrial Revolution
SDTr	Surrogate Digital Triplet
CMDTr	Condition Monitoring Digital Triplet
DTr	Digital Triplet
DT	Digital Twins
CPS	Cyber-Physical Systems
RCM	Reliability-Centered Maintenance
MSD	Mass-Spring-Damper
DAC	Digital to Analog Converter
MCSA	Motor Current Signature Analysis
AE	Acoustic Emission
FFT	Fast Fourier Transform

HHT	Hilbert Huang Transform
WT	Wavelet Transform
EMD	Empirical Mode Decomposition
IMF	Intrinsic Mode Functions
Hz	Hertz
kS/s	kilo-Samples per second
CWT	Continuous Wavelet Transform

Chapter 1

Introduction

There have been four industrial revolutions since the mid-1700s. The First Industrial Revolution (1IR) ranged from 1750 to 1840 and characterized changes in prior human and animal labor to early manual machinery [8]. Novel inventions during this era are the steam engine, the spinning jenny, the Bessemer process for steel manufacturing, and improved chemical production for manufacturing. The Second Industrial Revolution (2IR) from 1820 - 1920 was characterized by more common electrical technologies, the internal combustion engine, and developments along the assembly line [9]. The Third Revolution (3IR) the focused on incorporating more electronics and automation to the assembly line, further automating production [10]. Currently, industry is experiencing its Fourth Industrial Revolution (4IR) [11] with some researchers suggesting that society is preparing to move into the Fifth Industrial Revolution (5IR) [12]. 4IR focuses on integrating new process technologies, such as 3D printing, new fuel sources for transportation, such as electrification and hydrogen, and integrating cyber-physical systems and high-performance computing for artificial intelligence in industrial applications. 5IR will allegedly focus on personalizing the cognitive and human intelligence systems from 4IR.

Whether the world is in the 4IR, transitioning from 4IR to 5IR, or in the 5IR, there is a general push for leveraging the massive amounts of available data from industrial processes to make informed decisions. Subsequently, these decisions, backed by newly available data, effectively impact daily operations to minimize costs and ensure quality operations. The massive amounts of available data, termed Big Data [13], require digital tools to process the data into information in a timely and effective manner. Cyber-Physical Systems (CPSs) [14] and Digital Twins (DTs)

[15], with their variations, help to transform the available data into information. These systems have been implemented in multiple industries, such as digital twins in health care for elders for diagnostics and prognostics [16], digital twins for processing data during disasters and crises [17], digital twin applications for safety and security in transportation systems [18], digital twins for product development and management [19], cyber-physical systems in manufacturing for control and product quality [14, 20], and cyber-physical systems for energy control [21].

However, these condition monitoring (CM) systems require reliable, large amounts of data to determine the diagnosis and prognosis for the equipment. Generally, the data collected from a manufacturing environment contains incomplete class representations or the labels attached do not correctly characterize the equipment health state or operation [22]. Incomplete class representation refers to datasets that do not contain enough data for certain classes, which then cause skewed results and incorrect diagnoses. Leukel *et al.* [23] found that 38% of 34 reviewed articles did not discuss the breakdown of data. The nonexistent data breakdown is further exacerbated by datasets with few failures, which made up a further 10% of articles. The lack of reporting data and failure data availability makes it difficult to reproduce results and transfer any meaningful information and knowledge gained to industry, where these systems can make a difference.

The lack of reliable data leads to challenges in implementing condition monitoring systems in industrial systems. Transfer learning, data augmentation, and new algorithm development have each been implemented to address the data quality problem in condition monitoring systems. However, their integration into the long-term condition monitoring system leads to varying success levels for improving health predictions for the monitored equipment. This research proposes a novel approach to the data quality problem in manufacturing using surrogate systems to generate the necessary training data for condition monitoring systems in industrial settings. The process alleviates the need for consistently accurate labeled data from the manufacturing environment to train the CM system for failure diagnosis and prognosis.

1.1   Goal

The overall goal is to improve bearing diagnostic and prognostic predictions for equipment health by utilizing surrogate systems to generate failure data representing production equipment failure, thereby providing training data for condition monitoring solutions without incurring the

4

cost for real world failure data. The overall process is labeled as the Surrogate Digital Triplet (SDTr) or Condition Monitoring Digital Triplet (CMDTr) system, which has the following elements: a real-world production system that provides training and monitoring data (the physical triplet), a surrogate system that provides representative training data to the digital triplet (surrogate triplet), and the relatability and augmentation analysis for the surrogate system and condition health monitoring analysis for the production system (digital triplet). The fully-trained digital triplet informs the manufacturer about the production equipment's health state, allowing the manufacturer to take a proactive approach in their maintenance scheduling. With a comprehensive training profile for manufacturing equipment and the ability to generate training data in parallel with the production system, the ability to train and retrain predictive analytic algorithms allows for more robust and effective implementation with knowledge of the possible failure modes for the production system before any equipment failure occurs during production.

1.2 Objectives

To test this novel process, bearings are chosen as the experimental system due to their criticality and use in rotational equipment. Rotational equipment is a primary focus for condition monitoring as these systems are found in many industries, such as transportation, defense, manufacturing, and energy. Hence, the objectives are generalized for other equipment; however, the validation for this work is only conducted with bearings. The first objective is to create a methodology for generating data in offline environments for training online condition monitoring systems. The method standardizes damage plans to recreate "life-like" damage realistically. The second objective determines the procedures and necessary knowledge for transferring and merging the generated data with production application data. The third objective focuses on how machine learning models perform when considering the augmented bearing data compared to the original bearing data from the test system.

1.3 Questions

Each research objective has a set of questions to help guide the work and answer fundamental questions related to condition monitoring systems for bearings.

i.) Research Objective 1: Surrogate System Failure Data Generation for Bearings – Purposeful Failure Methodology

   a.) Research Question 1.1: How are the existing bearing datasets deficient for training manufacturing condition monitoring systems?

   b.) Research Question 1.2: What methodology criteria are necessary to fill gaps for condition monitoring datasets utilizing a surrogate triplet?

   c.) Research Question 1.3: How much data are needed to train a bearing condition monitoring system for diverse failure modes?

ii.) Research Objective 2: Data-Driven and Physics-Based Methods for Data Relatability & Transfer

   a.) Research Question 2.1: How is the artificially generated data from the surrogate system related to the production data taken from the real-world system?

   b.) Research Question 2.2: What considerations or knowledge are needed to relate different systems?

   c.) Research Question 2.3: Are the physical scaling and spectral augmentation sufficient for translating data from the laboratory environment to the real world when considering similar applications?

iii.) Research Objective 3: Evaluation of Condition Monitoring System for Multi-Fault Diagnosis - Bearings

   a.) Research Question 3.1: Is there an ideal amount of real versus augmented failure data for deploying offline data to a bearing diagnostic system?

   b.) Research Question 3.2: What augmented data process (data augmentation or physics scaling) increases the model performance regarding the assigned algorithm metrics?

   c.) Research Question 3.3: Does feature window optimization for bearing features improve bearing fault diagnosis regarding the assigned algorithm metrics?

1.4   Scope and Tasks

The proposed process entails the creation of a "Digital Triplet", a system consisting of a physical triplet, a surrogate triplet, and a digital triplet. The tasks are assigned based on the research objectives and explore the relationships of the physical triplet to the surrogate triplet (Research Objective 1), the surrogate triplet to the digital triplet (Research Objective 2), and the digital triplet to the physical triplet, where the monitoring is trained based on data from the surrogate triplet (Research Objective 3). The research objectives and questions are broken down into a detailed research scope with the corresponding tasks.

i.) Research Objective 1: Surrogate System Failure Data Generation for Bearings – Purposeful Failure Methodology

   a.) Research Question 1.1

   b.) Research Question 1.2

      1.) Task 1: Conduct a dataset assessment to determine the criteria for condition monitoring datasets. Evaluate prior condition monitoring and design methods.

      2.) Task 2: Formalize tasks into the Purposeful Failure Methodology (PFailM), which is compatible between component, system, production line, and factory applications.

   c.) Research Question 1.3

      3.) Task 3: Generate and characterize failure data from multiple failure modes to create a comprehensive failure profile for the surrogate system.

         I.) Generate Fatigue Damage Data

         II.) Generate Contamination Damage Data

         III.) Generate Brinelling Damage Data

         IV.) Generate Lubrication Damage Data

      4.) Task 4: Assess failure data using data-driven and physics-based equations associated with the equipment to determine the separability of data concerning the contrived failure modes.

ii.) Research Objective 2: Data-Driven and Physics-Based Methods for Data Relatability & Transfer - Bearings

a.) Research Question 2.1

    1.) Task 1: Conduct a comprehensive literature review of existing techniques to determine the best practices and limitations for the data transferability from a scaled system to a full-size system.

b.) Research Question 2.2

    2.) Task 2: Augment data with damage coefficients related to speed, load, and contamination level using spectral subtraction across bearings of different sizes.

    3.) Task 3: Perform relatability analysis relating bearing geometry and dynamic parameters between bearings of similar sizes in the surrogate system data using mass-spring-damper models.

    4.) Task 4: Evaluate the physical variations when considering systems under different operating conditions.

c.) Research Question 2.3

    5.) Task 5: Validate the system methodology by relating augmented data from the surrogate test stand to data from a real-world production system.

iii.) Research Objective 3: Evaluation of the Condition Monitoring System - Bearings

a.) Research Question 3.1

b.) Research Question 3.2

    1.) Task 1: Train and evaluate machine learning models under different mixtures of real and augmented failure data to diagnose bearing failure

        I.) Random Forest Classification

        II.) Decision Trees Classification

        III.) Naive Bayes Classification

        IV.) Nearest Neighbors Classification

        V.) Support Vector Machine Classification

        VI.) Multilayer Perceptron

        VI.) Deep Multilayer Perceptron

    2.) Task 2: Train and evaluate machine learning models under different mixtures of real and augmented failure data to provide a prognosis for impending bearing failure

I.) LSTM

II.) Hidden Markov Model

III.) Monte Carlo Markov Model

c.) Research Question 3.3

    3.) Task 3: Recreate the algorithms based on optimized time windows to determine if accuracy increases with different time windows.

Figure 1.1 outlines the work overall pathway. The dissertation work follows a linear pathway beginning with Research Objective/Area 1. Each research question (RQ) is addressed in the corresponding research area before moving onto the the following research objective. The work conducted in RQ 1.2 does impact RQ 2.1, and the work conducted in RQ 1.3 impacts RQ 3.1. The overlaps are addressed in the corresponding RQ sections.

Figure 1.1: PhD research outline: Goals, Objective Areas, Questions

Research Goal: Improve bearing diagnostic/prognostic performance by supplementing the training dataset with bearing data created on a separate application

Research Objective 1: Surrogate System Failure Data Generation for Bearings – *Purposeful Failure Methodology*

Research Question 1.1: How are the existing bearing datasets deficient for training manufacturing condition monitoring systems?

Research Question 1.2: What methodology criteria are necessary to fill gaps for condition monitoring datasets utilizing a surrogate triplet?

Research Question 1.3: How much data are needed to train a condition monitoring system?

Research Objective 2: Data-Driven and Physics-Based Methods for Data Relatability & Transfer

Research Question 2.1: How is the artificially generated data from the surrogate system related to the production data taken from the real-world system?

Research Question 2.2: What considerations or knowledge are needed to relate different systems?

Research Question 2.3: Are the physical scaling and spectral augmentation sufficient for translating data from the laboratory environment to the real world when considering similar applications?

Research Objective 3: Evaluation of Condition Monitoring System for Multi-Fault Diagnosis – Bearings

Research Question 3.1: Is there an ideal amount of real versus augmented failure data for deploying offline data to a bearing diagnostic system?

Research Question 3.2: What augmented data process (data augmentation or physics scaling) increases the model performance regarding the assigned algorithm metrics?

Research Question 3.3: Does feature window optimization for bearing features improve bearing fault diagnosis regarding the assigned algorithm metrics?

Chapter 2

Background

The following chapter is considered an introduction to the relevant background sections for this work. These sections provide an overview of background work related to condition monitoring system and highlight the relation to the research goal. Section 2.1 provides a brief overview of bearings. Section 2.2 and 2.3 provide motivations for the research work related to maintenance costs and work strategy. Section 2.4 through 2.6 describes the different data acquisition, processing, analysis, and manipulations investigated. Section 2.7 introduces technologies that package and interface between the digital and physical systems. Figure 2.1 maps how the background topics relate to different portions of the dissertation. Additional background information may be provided in each of the separate research objectives called out. These initial background sections are meant to provide the basis for further discussion in a deeper description of the research work in Section 3.3.
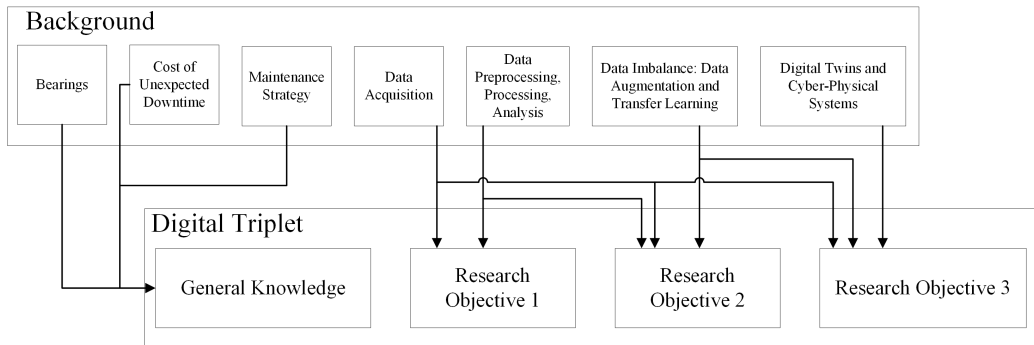
Figure 2.1: Path to show how the background topics relate further to the research work as described in the next section

2.1 Bearings

Bearings are critical in rotational equipment as they help transfer load and motion between different systems. The work in this dissertation focuses explicitly on rolling element bearings. Four major parts define rolling element bearings. The first two are the inner and outer rings found in other types of bearings. The difference comes from the other two parts, the rolling element and the cage. In plain bearings (*i.e.*, bearings with no rolling element), the inner and outer rings are separated by a film of grease. In a rolling element bearing, the rings have grooved trenches called raceways. Rolling elements are placed in the raceway and rotate with the bearing as it moves. The rolling elements distribute the load between the two rings. A cage holds these rolling elements to ensure equal distribution around the bearing.

With these four parts, multiple configurations for rolling element bearings depend on the application. For example, a deep groove ball bearing is suitable for all-around applications. However, due to a different configuration, a thrust bearing would perform better in a purely axial consideration but could not handle significant loads in the radial direction. Figure 2.2 shows an example of a deep groove ball bearing in (a) and a thrust ball bearing in (b). Any significant radial load applied to the thrust ball bearing could cause early failure. In these two examples, the rolling elements are balls; however, they can also be cylindrical in other applications. The cylindrical applications accommodate the scenarios of pure radial that might occur in certain applications. There is a third type of spherical roller, which would accommodate misalignment scenarios. Misalignment is where the shaft's geometric center line differs from the bearing's center line. Due to their precise nature, small amounts of misalignment can cause early damage to bearings, causing them to fail faster. For bearing selection, major bearing manufacturers (*e.g.* SKF and Koyo/JTEKT) provide processes to select bearings for the correct application based on the application requirements, operating conditions, lubrication, operating temperature and size.
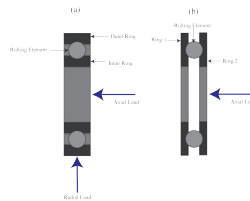
Figure 2.2: Two bearing types: (a) Deep Groove Ball Bearing and (b) Thrust Bearing

Further documentation is available from the American Bearing Manufacturers Association (ABMA) and the International Organization for Standardization (ISO). These two organizations standardize bearing dimensions and coding to ensure that bearing dimensions are identical for every bearing manufactured. ABMA standardizes bearings using the English system, and ISO standardizes bearings using the metric system. During this research work, bearings are defined by the metric system. The primary bearings used in this dissertation work are deep groove ball bearings, denoted by the convention 6XYY, where X represents the width series and YY codes for the bore size. Due to their simplistic but precise nature, these bearings are considered a good all-around fit for most applications, as they can accommodate radial and axial loads. Typical applications in industrial environments for deep groove ball bearings involve motors, gearboxes, pumps, and fans. Deep groove ball bearings are also found in many home applications, including air conditioners, washing machines, and vacuums. Because of their prevalent nature, this research focuses on determining the most effective manner to train, deploy, and execute condition monitoring systems to reduce unexpected equipment failure.

2.2 Cost of Unexpected Downtime: Bearings

2.2.1 General Unexpected Downtime Cost

For this work, unexpected downtime is defined as any incurred downtime not accounted for in manufacturing operations. Another term for labeling unexpected downtime is simply as operation disruptions. Veira *et al.* [24] considered the following as disruptions that caused the need for manufacturing rescheduling: machine failure, Urgent jobs, job cancellation, change in delivery dates, delay in materials, change in job priority, rework issues, misallocation over process time, and operator mishaps. As seen with the recent COVID-19 pandemic, disruptions in the timing for more materials and products impacted the revenue streams for all industrial enterprises [25, 26]. Certain strategies, such as supply chain redundancy, localization, and digitalization, improved supply chains to respond to the supply chain changes forced on companies due to the pandemic. Adopting these new strategies allows supply chain adaptability when presented with a set of adverse conditions.

However, machine failure is the most cited manufacturing disruption outside supply chain disruptions by Vieira *et al.* [24]. While the primary focus is machine failure in manufacturing for this work, machine failure causes disruptions in every industry, such as transportation [27], healthcare

13

[28], and energy [29]. Equipment failure in each industry incurs additional costs outside traditional monetary values and product loss. For example, equipment failure in the healthcare industry can lead to patient deaths. Heneghan *et al.* [30] found that 44% of the medical device alerts issued over five years had a reasonable probability of causing severe health consequences or death due to defects found in their medical devices. Another instance is the failure of hospital systems like air conditioning. While critical for climate control, air condition failures impact hospitals to filter air, reducing their ability to fight infection while treating people [31]. Failures in transportation could cause additional delays that strain other services. Gerdes *et al.* [32] found that air condition system failures caused approximately $5.50 per flight. Gerdes *et al.* acknowledged that value may be low and the cost savings from CBM implementation were only $1.1. However, when combining those numbers with the approximately 16 million flights per year [33], the overall cost of just the air conditioning failures represents a more significant cost.

Critical equipment failure in manufacturing has a degree of cascading effects, depending on the industry, the equipment, and the failure type. The worst case scenario in any equipment failure is loss of life, which is unquantifiable in cost. However, Chong *et al.* [34] found that facilities management and maintenance teams prioritized safety to reduce the possibility of injury and potential death. With this consideration, the next worse case scenario, in terms of cost, is a failure that causes a production stoppage. Thomas *et al.* [1] found that the approximate maintenance expenditures for NAICS 32 (Forestry), 332 (Metal Fabrication), 333 (Machinery Manufacturing), 334 (Computer and Electronic Product Manufacturing), 335 (Electrical Equipment, Appliance, and Component Manufacturing), 336 (Transportation Equipment Manufacturing), and 339 (Miscellaneous Manufacturing) were estimated as $74.5 billion nationally in 2016. The cost breakdown from that number was divided into expected expenditures at $57.3 billion, the additional costs incurred by faults and failures at $16.3 billion, and the inventory cost at $0.9 billion. In addition to these costs, losses related to downtime, defects, and delays were $119.1 billion. The amount of injuries each year related to equipment failure was 134.1 injuries and 0.4 deaths on average for maintenance issues. Table 2.1 displays these numbers with their 90% confidence intervals. The survey and analysis were based on 71 companies from the referenced NAICS codes.

The cost analysis provides a general overview of costs affecting the manufacturing industry across the nation. However, to tailor the price to particular companies, the economic cost function changes based on the company size, the industry, and the failure type. For example, Muller *et al.* [35]

14

Table 2.1: Table from Thomas *et al.* [1] showing the breakdown in costs and losses within their 90% confidence interval

Costs and Losses	Estimate ($ billion)	Confidence Interval (90%) ($ billion)
Direct Maintenance Costs	57.3	42.4 - 72.2
Costs due to Faults and Failures	16.3	7.1 - 25.2
Inventory Costs	0.9	1.3 - 5.6
Unplanned Downtime Losses from labor	13.5	7.1 - 22.1
Unplanned Downtime Losses from Building Depreciation	2.5	1.8 - 3.1
Unplanned Downtime Losses from Machinery Depreciation	1.0	0.7 - 1.2
Unplanned Downtime Losses from Energy	1.1	0.8 - 1.4
Defect Losses	0.8	0.0 - 2.7
Lost Sales from Defects	31.2	3.6 - 58.7
Lost Sales from Delays	69	29.8 - 108.1

found that an hour of motor downtime could incur between $7,000 - $200,000 costs, depending on the industry. Another example considers the downtime cost for automotive manufacturers as $ 22,000 per minute [36]. With this wide range of cost, what is the cost of a bearing failure in manufacturing? The section purpose was to point out not so much the cost for the bearing replacement but more so the cascading costs associated with that bearing failure. A 6205 bearing may cost $30 to replace typically, but if it fails within side a motor and is not assessed properly, it could cost tens of thousands of dollars worth of damage to that motor.

2.2.2   Unexpected Downtime Cost: Bearings

The bearing failure rate for systems in the industry varies based on that particular system. SKF, a major bearing manufacturer, estimated that 1 billion of the 10 billion bearings manufactured yearly are replaced. From that 1 billion, approximately 50 million bearings are replaced each year due to a failure [7]. For example, in industrial and commercial motor applications, Rao *et al.* [37] estimated 14 million hours of downtime. For those failures, bearings were one of the number one leading maintenance issues at 7%. In a survey in the 1980s for industrial motors, O'Donnel *et al.* [38, 39, 40] published a three-part motor survey that found bearings contributing up to as much as 50% of the failures they investigated. There is a wide variance between the number of failures caused by bearings. That variance could be due to the sample selection and industries surveyed. Other industrial applications focus on bearings as possible failures, such as computer numerical controlled (CNC) cutting tools, which are susceptible to thermal seizure with their bearings [41], wind turbine gearboxes where the bearings are susceptible to axial cracking [42], and pumps where the bearings

fail due to high speed and poor lubrication [43].

Bearings remain a critical component outside industrial applications, and standard failure mode requires extensive monitoring. For example, helicopters are another essential system that sees several bearing failures. Davies *et al.* investigated data obtained from the AgustaWestland Limited Materials Technology Laboratory that found that the failure investigations from the lab found that 55% of helicopters experienced some form of fatigue failure. Within this 55%, bearings contributed the second most failures at 17% after gears at 20%. Prizinger *et al.* [44] conducted a similar survey with U.S. Army helicopters to identify primary failure modes and provide tools for CBM. They found that the main rotor swashplate (MRSP) and the tail rotor driveshaft (TRDS) bearings or bushings contributed 36% and 23% of failures, respectively. These two systems are critical to aircraft control. If the MRSP bearing fails, the control capability to the main rotor pitch links is reduced. If the TRDS bearings fail, the pilot loses anti-torque and directional control of the aircraft. There is also a significant time component to replace and repair these bearings. For the MRS, the maintenance hours can extend up to 29.2 hours, whereas 11.5 hours are expended with replacing both the TRDS hanger bearings (forward and aft).

The point of the preceding paragraphs is to highlight how critical bearings are in equipment. Not only that, but depending on the location, the industry, the application, and the bearing type, the repair time and cost will vary heavily. On average, planned and unplanned maintenance costs can cover up to 40% of manufacturing costs [45]. These costs are derived from factors such as the size of the application, the distance to transport parts and personnel to perform repairs, the new equipment costs, the storage costs, the labor cost associated with the repair, the labor cost from idle operators, and any lost sales. For example, in offshore wind turbines, Walford [46] found that the size of the wind turbine drove the high costs. Replacing a gearbox in a 660 kW turbine on a 65-meter tower was $120 thousand. 80% was consigned for the gearbox, with the rest detailed for the actual repair at the site. Increasing the turbine to a 1.5 MW turbine on an 80-meter tower increases the gearbox cost by 3 or 4 times. With gearbox failures in wind turbines, Sheng [47] found that approximately 70% of gearbox failures stemmed from bearings.

Maintenance costs vary from organization to organization due to the industry's structure and how the organization operates. Hence, instead of directly quantifying costs, Figure 2.3 and 2.4 shows two parts of a maintenance cost flow chart for motor bearing failures. Figure 2.3 shows the direct failure costs immediately after it occurred, whereas Figure 2.4 focuses on the repair and further

16

failure effects after the repair. The terminology in this flow chart applies to other manufacturing systems; the motor bearing failure is merely used as an example. After a failure has been identified, the first action is determining how it affects production. A motor failure on the production line would imply that the motor drives a system that either moves the product from station to station or actuates a process to act on the product. A motor failure not occurring on the production line may correspond to a support center not directly tied to production. These failures do not hamper production but may remain a critical cost as specific tasks may have to be outsourced.
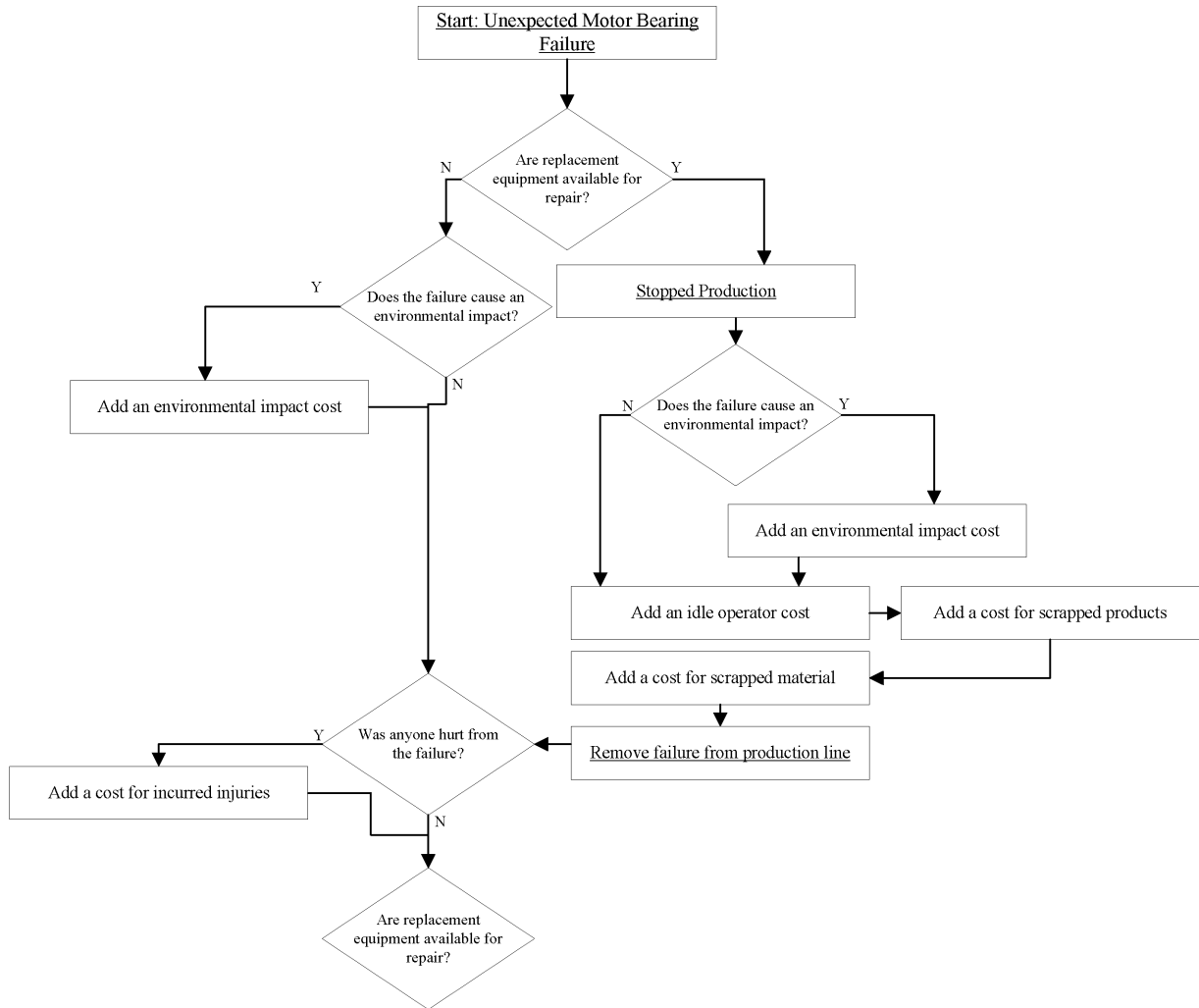
Figure 2.3: A flowchart for considering the cost of failure (Top Portion of the flow chart. Split to accommodate two pages)

In stopped production costs related to unexpected downtime, there are five costs to consider: environmental impact, idle operator costs, scrapped material, scrapped products, and incurred in-

17

juries. Environmental impact costs from failures occur when harmful substances are released into the environment due to equipment failure. These cases occur all the time in manufacturing due to improper controls or technologies employed to mitigate the release of pollutants. For example, Chevron Phillips Chemical Company was ordered to pay $3.4 million [48] due to violating the Clean Air Act in the United States. On February 3, 2023, a Norfolk Southern train derailed due to a faulty bearing. The EPA has ordered Norfolk Southern to clean up the site, costing approximately $70 thousand a day [49]. In addition, there have been recent calls for Norfolk Southern to pay for the health costs due to the environmental impact [50]. Environmental costs are incurred if the motor failure occurs within a system designed to prevent this release. When production stops, operators still need to be paid for their present time. The "idle" labor cost is avoidable during scheduled shutdowns as manufacturers give time off to operators. Scrapped material and production costs include additional costs outside of direct loss. For replacing material scrap, there is a cost for purchasing new material, transporting the new material, and preparing the material for production. Kalpajian *et al.* [51] found that material costs could account for as much as 50% of costs in modern manufacturing. For a scrapped product, those costs further include the delay and defect costs to customers. The delays and defects from equipment failures caused the most significant losses from the cost analysis of Thomas *et al.* [1]. These subsequent losses could reflect a trust cost or a further unreliability cost attached to the manufacturer. It may prevent one firm from returning to do business with another firm and affect a company's ability to remain competitive

These cost areas reflect only the costs that occur with stalled production. Further costs are incurred from the damage and repair assessment for the failure. For instance, injury costs are added if someone gets hurt due to an equipment failure. Current manufacturing systems are designed to account for safety and risk; however, injuries can still occur on the manufacturing line. Dunning *et al.* [52] found a significant cost for medical claims related to musculoskeletal injuries in Ohio manufacturing at $673 million. Manufacturing was only second to the service industry, with a cost of $909 million. Not all of these are attributable to equipment failure; however, the probability of adding additional cost due to equipment failure is higher.

There are typically two scenarios to consider for the actual equipment repair cost: replacement or no replacement. If there is a replacement, an inventory cost must be added that incorporates a carrying or shelf life cost. There is also a cost related to maintaining the replacement and ensuring it is still operable when needed. After preparing the replacement, the repair work can begin and

18

include both in-house and outsourced labor. Outsourced maintenance help would comprise specialists to assist with specific repairs like equipment balancing and validation or used as storage for replacement equipment [53]. Motor specialists may become involved if the failure is due to a system misalignment, balancing, or looseness to ensure any installation does not affect other equipment.
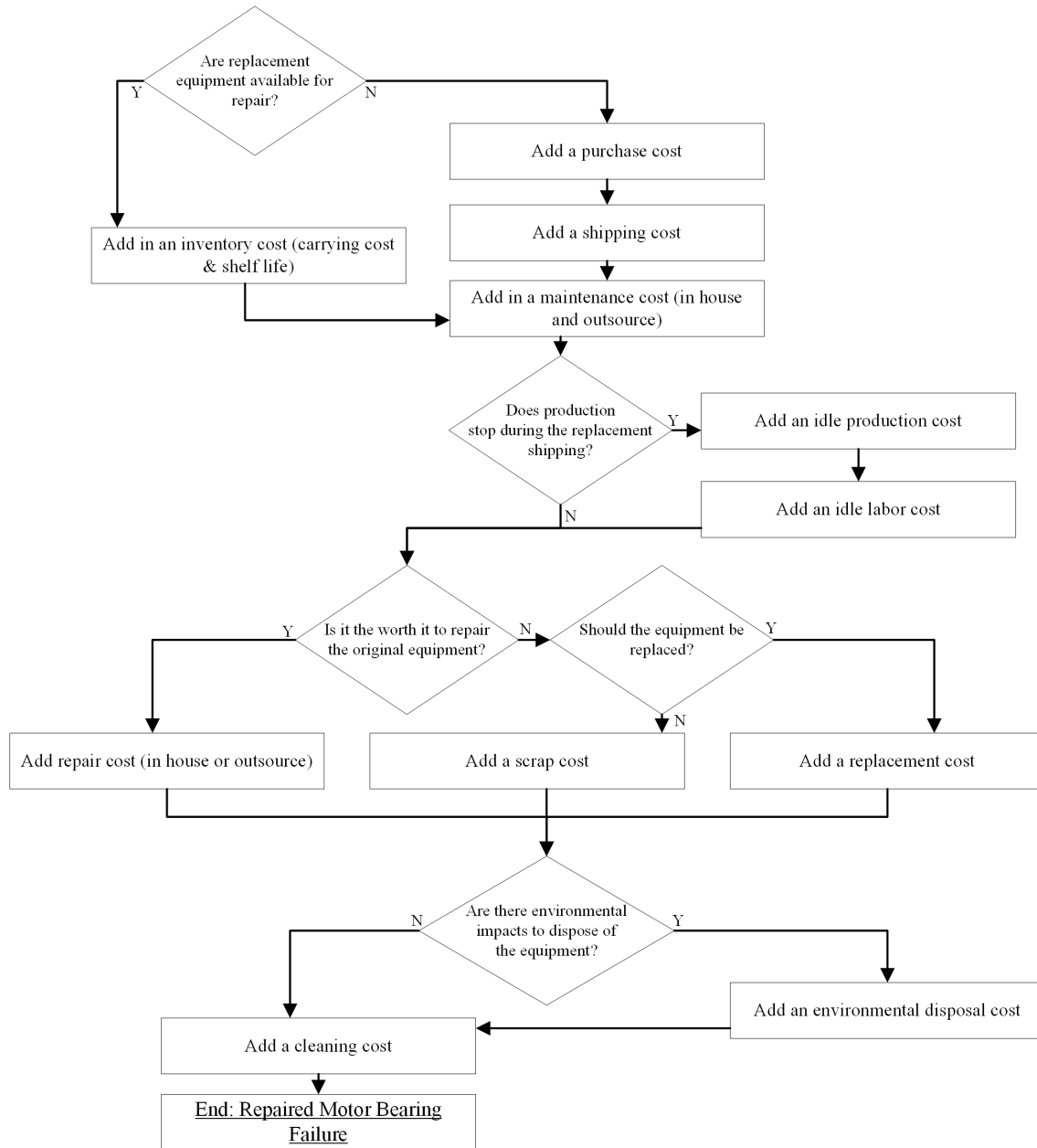
Figure 2.4: A flowchart for considering the cost of failure (Bottom Portion of the flow chart. Split to accommodate two pages)

If replacement is unavailable, additional costs may be incurred due to receiving new equipment. Purchasing and shipping costs are included in receiving new equipment. Depending on the motor size, it may come directly from the manufacturer rather than a supplier. For other equipment, additional costs are incurred from purchasing any new tooling or services to run the equipment. These are in addition to any costs needed to perform the repair, either with in-house specialists or bringing in outside experts. In addition to the services performed, an additional cost consideration for any production delays resulting from waiting for the repair. Repairs are not immediate, even when they don't affect the production line. Equipment that is down for a prolonged period could affect the production line indirectly and force a production shutdown.

The next consideration is whether it is worth it to repair the original equipment. In the case of the wind turbine gearboxes investigated by Walford [46], these systems are large enough scale that repair can return them to a "like new" or similar condition. The repaired components can then be moved to storage or placed back on the production line at the following scheduled downtime. If the equipment is not repaired, a replacement may be purchased, which will incur an additional cost. If a replacement and repair is not in the maintenance policy, a scrap cost related to any disposal must be assessed. A scrap cost would involve transporting and disposing of the equipment. Furthermore, the disposal should be checked for any subsequent environmental considerations in the disposal. Finally, a cleaning cost is assessed for the equipment operating area to reduce the possibility of contamination or early failures.

A large amount of this cost is incurred due to improper planning caused by untimely failure. The cost is less substantial by implementing condition monitoring, especially for bearings. Bearings, by themselves, are a much lower cost for repair and replacement than a complete system failure due to a bearing failing. Bearing CM allows maintenance teams to schedule the failure time to coincide with the time that does not affect production. Proper maintenance scheduling reduces the potential environmental impact, production costs and the lead time to wait for repair equipment. In an ideal scenario, none of the production costs from Figure 2.3 and 2.4 are included in the final cost for maintenance.

2.3 Maintenance Strategy

Different strategies are employed based on the identified costs in Section 2.2.2. Cooke [54] investigated plant maintenance strategies for different British manufacturing firms. They determined that maintenance changed as a result of generational changes. The first generation, identified before the Second World War, adopted the maintenance approach of "fix it when it is broken". This maintenance approach has also been termed corrective maintenance [55]. Maintenance fixes equipment as it breaks on the production line. The second generation, identified from Cooke [54], comes from the 1970s. The second generation focused on implementing more preventive measures, giving rise to the practice of Preventive Maintenance (PM). The primary cause for this shift in maintenance stemmed from the use of more complex systems in manufacturing.

PM is implementing maintenance checks on a timely schedule based on the equipment health [56]. During PM implementation, maintenance schedules are based on the equipment life concerning a hazard function. The maintenance actions that affect the model are typically characterized as returning the equipment to either a perfect, imperfect, or minimal repair. A perfect repair returns the equipment to a near "new" state. An imperfect repair returns the maintenance to some level of functionality and restores a measure of life depending on the action. Finally, the minimal maintenance action implies the repair was only to ensure the equipment can continue to operate. Models for PM are framed using the time to failure to describe the equipment survival probability relative to after a repair is performed. Cranfield [57] incorporated Weibull distribution parameters to estimate the changing failure rate based on nonperfect and unknown maintenance actions. The model and approach employed by Cranfield are part of age reduction models, which use a hazard function to describe the system [58]. Two other model categories are hazard rate models and hybrid models. Hazard rate models consider the possibility of changing the hazard rate due to a dynamic environment. Xiaofei *et al.* [59] considered a changing hazard rate function to accommodate a changing environmental state. Hybrid models consider changes to the hazard function and rate as equipment life fluctuates. Yang *et al.* [60] modeled shock damage as a change to the hazard function and rate, replicating the abrupt degradation and drop in equipment life.

After the second generation, Cooke [54] describes the third generation of maintenance as focusing on more condition monitoring and reliability-centered maintenance. Reliability-Centered Maintenance (RCM) and CBM are designated under PdM in a tree of maintenance strategies [61].

Niu *et al.* [62], however, considered CBM under PM strategies. Between PM and PdM, the line begins to blur between maintenance strategies. Even further blurring the line, Niu *et al.* [62] combines CBM and RCM to optimize maintenance costs. The separation between PM and PdM is that PM focuses on time checks between hazard functions and rates, whereas PdM focuses on scheduling maintenance checks before failure. PdM uses models based on the data collected from the equipment to corroborate with the expected lifetime during operation [63]. Based on the data collection, these models could be data-driven, physics-based or knowledge-based. A data-driven model considers changes based on the measured data and associated changes within a system. Deep and machine learning algorithms are considered one form of data-driven modeling to detect changes in the collected data [64]. Physics-based models use the measured data to determine a change in the system health state using a physics equation [65]. Robot predictive maintenance incorporates robot physics to validate expected paths with physical data [66]. Deviations in robot paths may correspond to degrading components that require replacement. Knowledge-based models consider the simulation of expert and domain knowledge combined with data-driven modeling. Cao *et al.* [67] create a knowledge-based approach based on statistical and symbolic modeling to extract degradation models and predict machine failures. The data extraction occurred using chronicle mining to incorporate contextual knowledge. Hybrid models are considered a combination of two or more models described before.

Figure 2.5 shows an example of the progression of maintenance strategies and the philosophy behind each. The top picture represents a failed engine from a piece of equipment. A photo shows that a failure has prevented the equipment from being operable. If indicators had been in place to catch the failure, it may have been detected earlier. However, the CM approach dictates that repair only occurs after failure. The middle picture corresponds to PM. In this case, oil checks occur regularly to ensure proper lubrication for the car engine. These time checks are based on the engine life and the lubrication type, similar to the time checks based on the equipment life in manufacturing. These periodic checks ensure that equipment remains operational over time. However, if there is a drastic shift in equipment life over time, failures could still occur between the predetermined maintenance checks. The remaining amount of oil is displayed on the dashboard in the bottom picture. For the PdM approach, there is now an indicator and model that changes based on the oil and equipment condition. The maintenance team monitors this value until a predetermined level when maintenance is finally conducted.

Corrective Maintenance
Prior to 1945

Preventive Maintenance
1970s

Predictive Maintenance
1980s

Figure 2.5: Examples of the different maintenance strategies

With the three iterations of maintenance and their examples shown in Figure 2.5, the costs for maintenance will change between each type. Figure 2.6 considers the maintenance costs concerning their maintenance activities. The cost comparisons are considered across three categories: total cost, prevention cost, and repair cost. The prevention cost corresponds to what is conducted to prevent failure. A repair cost corresponds to what is conducted to fix a failure. CM considers a high repair cost but a low prevention cost. PM considers a high prevention cost and low repair cost. PdM, in the optimal state, considers an equal amount of repair and prevention costs. PdM
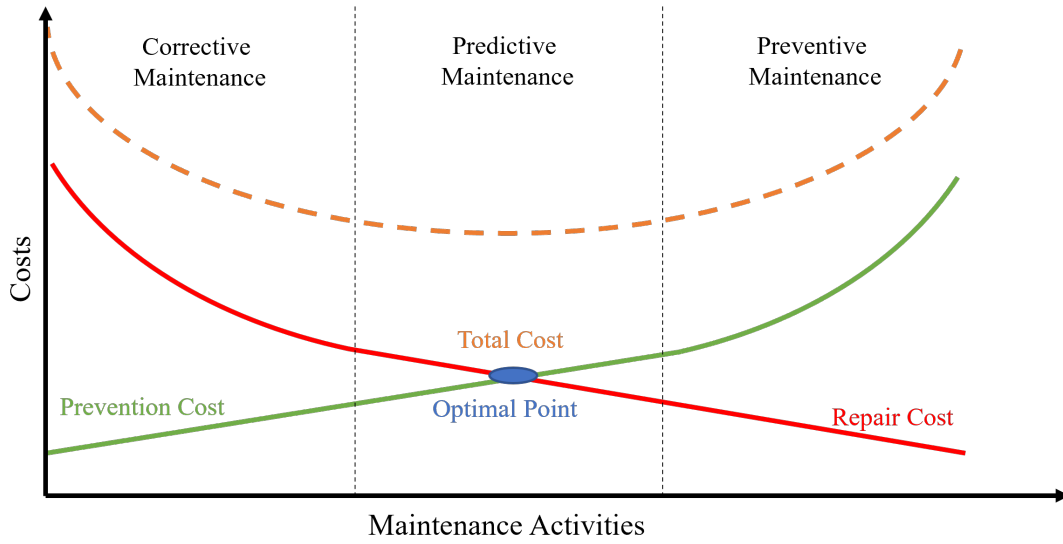
Figure 2.6: Total Cost Strategy

adoption brings in additional avenues of data collection for information generation, reducing the maintenance checks that occur with PM. The adoption reduces the potential of a high-risk failure that could occur with CM. The cost reductions of accepting PdM vary based on the equipment and industry. A list of several references to cost savings related to predictive maintenance is below:

i.) Thomas *et al.* [1] found that the adoption of predictive maintenance decreased downtime by $6.5 billion and an increased sales of $67.3 billion.

ii.) Adu-Amankwa *et al.* [68] predicted an average cost savings ranged from £22 thousand to £48 thousand when using predictive maintenance.

iii.) de Pater *et al.* [69] conceived a maintenance strategy using prognostic modeling for turbofan engines to reduce engine failures using task rescheduling dynamically. After the rescheduling, turbofans only contributed to 7.4% of total maintenance costs over 31.8%.

iv.) Umeda *et al.* [70] reduced maintenance costs to 69% on average for component failures in plasma etchers based on using RUl prediction methods.

2.4   Data Acquisition

Data acquisition in condition monitoring considers the different methods for collecting data from critical equipment. The data comes from integrated sources during the equipment design or from later data-gathering devices added to the equipment. Data acquisition is often seen as the first step for condition monitoring [71]. The collected data are then converted into data features for analysis and information generation, which is described in Section 2.5.

Due to the need for large amounts of verifiable data, standards exist to gather these data from condition monitoring equipment. Table A.1 contains all the condition monitoring standards from ISO TC 108/ Section 5. TC stands for technical committee. ISO TC 108 is the committee for mechanical vibration, shock and condition monitoring standards. Subcommittee 5 of this TC specifically focuses on standards related to the CM and diagnostics of the system. The general standards for vibration condition monitoring were included in Subcommittee 2. Their primary focus is measuring and evaluating mechanical vibration and shock applied to machines, vehicles, and structures. Hence, the vibration condition monitoring standards are in a different committee, than the other CM standards. The following subsection comprises the various methods for condition monitoring derived from research and the general concepts in detecting failure in equipment.

The following subsections consider vibration, electrical current, acoustics, temperature, and ultrasonics data acquisition methods concerning bearings. Another method that needs to be covered in depth is oil wear analysis [72], which counts the number of contaminating particles in the oil. However, the method is heavily subjective to the oil sample taken for the analysis. The method is commonly paired with other condition monitoring techniques to verify readings from these analysis techniques. Each subsection provides a brief overview of the CM strategy and the configuration. The following section (Section 2.5) describes more of the different processing methods for bearing condition monitoring.

2.4.1   Vibration

Vibration is the repetitive motion about an equilibrium point from a mechanical system [73]. Generally, vibration comprises two different components: frequency and amplitude. Vibration amplitude is the maximum distance, velocity, or acceleration reached over time and frequency. The frequency is the rate for how many cycles appear within a set period (usually one second). Generally,

vibration can be further defined into three categories: free vibration, forced vibration, and damped vibration [74]. These categories are usually defined using system dynamic principles surrounding mass-spring-damper (MSD) systems. Equation 2.1 represents the general equation for an MSD system. Each part of the MSD represents a different component of vibration that then equals the dynamic nature of the system, usually Force, denoted by $F(t)$ (N or lbf). The mass (m) coupled with acceleration ($a(t)$) refers to the inertial elements of the system. The spring elements (k) coupled with displacement ($d(t)$)) correspond to the stiffness within a system. The damper elements (b) coupled with velocity ($v(t)$) comprise mechanical resistance elements. In vibration, these are the parts of the system that dissipate the system energy.

$$F(t) = ma(t) + bv(t) + kd(t) \tag{2.1}$$

Table 2.2 considers the four possible vibration scenarios based on the three different categories of vibration. Free vibration is categorized as not having an acting outside force on the system. Depending on the scenario, there may or may not be a damping element as demonstrated by Equations 2.2 and 2.3. In both cases, the vibration occurs from an initial input. However, in the undamped case, the vibration continues until acted upon by either a damping element or an outside force. For the free and undamped vibration, the system eventually comes to rest. Forced vibration is characterized by an external force acting on the system. The force can vary with time and, in some cases, is represented by inertial, spring, or damping elements. Equations 2.4 and 2.5 demonstrate the Forced vibrations with and without damping, respectively.

For bearing condition monitoring, vibration is the standard method for detecting changes in condition. Vibration can be measured using displacement probes, velocity transducers, and accelerometers [75]. In an ideal configuration, three sensors are used to measure the axial, horizontal, and vertical radial vibration based on industry standards from ISO in Table A.1. Figure 2.7 shows the general configuration when considering three vibration measurement locations. In application, though, the radial location is difficult to reach safely during equipment operation; hence, most arrangements only consider the radial sensors. Using the two radial sensors, it is possible to sync the data from both sensors and conduct a phase analysis of the bearing vibration samples [76]. A system fault is apparent if the bearing signals are not 90°apart from each other. The vibration sensors should be positioned directly over the bearing and directly mounted to the housing using a

Table 2.2: Four vibration scenarios based on the three vibration categories

Vibration Category	Elements	General Equation	
Free Vibration without Damping	Inertial (m), Spring (k)	$$ma(t) + k(x(t)) = 0$$	(2.2)
Free Vibration with Damping	Inertial (m), Spring (k), Damping (b)	$$ma(t) + b(v(t)) + k(x(t)) = 0$$	(2.3)
Forced Vibration without Damping	Inertial (m), Spring (k), Force $(F(t))$	$$ma(t) + k(x(t)) = F(t)$$	(2.4)
Forced Vibration with Damping	Inertial (m), Spring (k), Damping (b), Force $(F(t))$	$$ma(t) + b(v(t)) + k(x(t)) = F(t)$$	(2.5)

fastened connection for the best signal. Other mounting methods are possible; however, the bearing frequency range degrades using other less rigid connections. Generally, the gathered vibration is passed to a signal conditioner and then digital to analog converter (DAC) before storing the data on a processing device, such as a microcontroller or computer. Vibration condition monitoring systems data can either pass through wired or wireless connections [77, 78]. The wireless connections depend on the operating environment, as specific wireless frequencies can disrupt production equipment.

2.4.2 Electrical Current

Electrical current condition monitoring identifies bearing faults in electro-mechanical systems by the electrical signals that power the system. The most common systems to use electrical condition monitoring are induction motors, where electrical current condition monitoring is termed Motor Current Signature Analysis (MCSA). MCSA monitors primarily the stator current. In an electrical motor, two primary components cause the equipment rotation: the rotor and the stator [79]. The stator is the stationary portion of the motor attached to the outer housing, while the rotor rotates around the center axis of the stator. The rotation comes from the magnetic field created by the electrical windings in the rotor and stator. The bearings are placed on the drive and fan end of
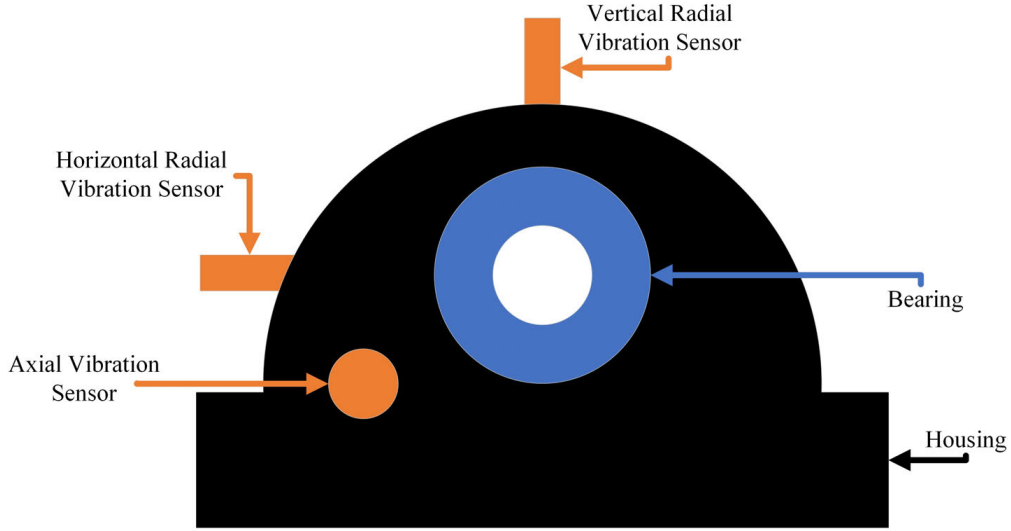
Figure 2.7: Bearing condition monitoring layout for a mounted bearing

the rotor and stator to help maintain alignment during rotation.

To perform current CM using MCSA, the sensors are placed on the electrical feed lines connected to the electrical system [80]. Current sensors usually employ the Hall effect to measure the electrical current for MCSA [81]. A Hall effect current sensor measures the magnetic field generated by the electrical current passing through the wire. The induced magnetic field is proportional to a voltage, which could be read to determine the current passing through the wire. Typically, a clamp-on sensor is used to measure the electrical current using these sensors. There are different types of sensors: open-loop and closed-loop sensors [82]. A closed loop sensor utilizes negative feedback with the hall element sensor to mitigate the linearity and gain errors associated with open loop sensors. However, due to the feedback from the system, stability is an issue where the system has difficulty responding to quick changes in a system. Figure 2.8 demonstrates the overall configuration for the hall sensor in the open loop configuration. A closed loop hall sensor considers an additional feedback line with a wire wrapping around a magnetic core.

Generally, induction motors are classified based on the number of phases that supply power. In industrial settings, a three-phase motor is commonly used due to the ability to provide a near constant alternating voltage supply to the motor poles to maintain instantaneous power [79]. The alternating voltage supply supports the magnetic field to cause the rotation. In a single-phase motor, the voltage supply varies uniformly between the different poles and cannot maintain the
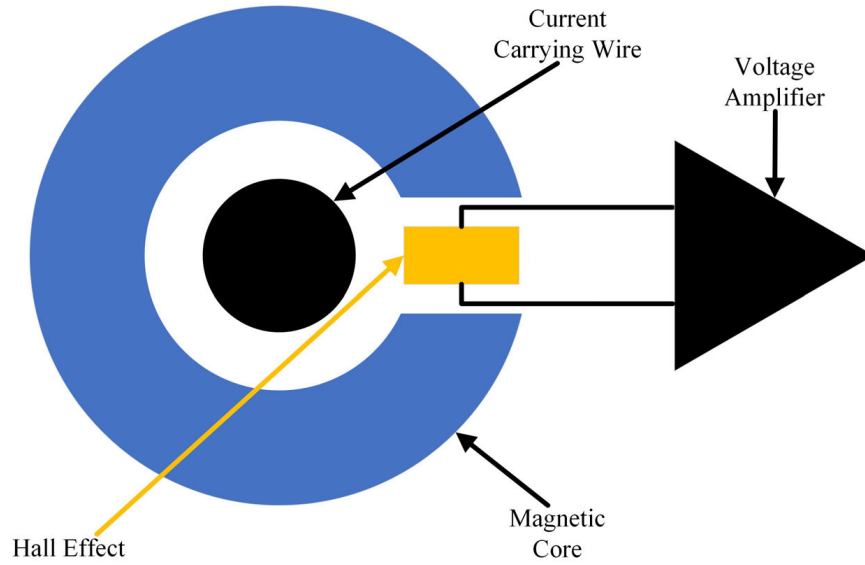
Figure 2.8: Open loop Hall sensor. A closed loop hall sensor has a feedback line connected to an additional wire wrapping around the magnetic core.

instantaneous power constantly. Three hall sensors are needed to wrap around the motor current supplies individually to monitor all three phases of an induction motor. The separate current sensors monitor a feed line input into the motor or electrical equipment. In the case of a single-phase motor, only one sensor would be needed on the "live" wire. Generally, these three sensors are connected to a signal conditioner and then DAC before storage and analysis. For electrical equipment, current monitoring is paired with vibration monitoring to provide increased accuracy for system predictions [83].

2.4.3 Acoustic Emission & Microphone Capture

Acoustic emissions (AE) are elastic waves generated by the rapid release of energy from a stressed material [84]. The phenomenon is primarily used in material fatigue detection in materials to determine the exact point at which a crack has occurred. In bearings, cracks appear due to increased material fatigue or thermal cracking. For the fatigue case, these cracks may appear in the subsurface and are only as easily detectable for vibration once it is too late in the case of vibration. Acoustic sensors are placed on the material to capture the best wave signature. Similarly to vibration sensors, AE sensors need to be placed close to the bearing to experience a clear signal. It has been found in structural analysis that sensors further away from the disturbance location
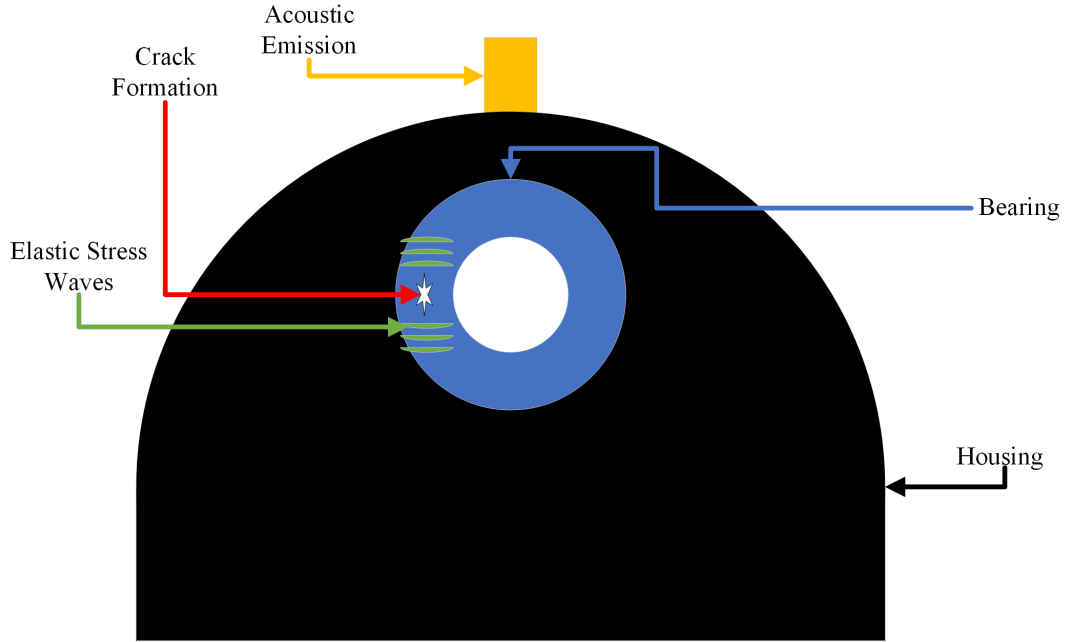
Figure 2.9: AE sensor with crack

elicit weaker signals [85]. However, the scale is different for steel structures and bearings. AE has been used extensively for monitoring changes in the lubrication [86] and contamination [87]. Figure 2.9 provides a rudimentary representation of AE sensing with a crack formation on the bearing. AE sensing is often paired with vibration CM to verify or enhance fault detection [86, 87, 88].

In addition to AE sensing, sound capture using microphones or other audio sensors is another possible method for analyzing bearing signals. An increase in noise generation from bearings and other rotational equipment implies metal-on-metal contact. The increase in contact implies that there is either inadequate lubrication or a potential defect on the the bearing elements. One method to monitor changes in sound is using microphones to measure sound pressure [89]. Sound pressure is the deviation in local pressure from ambient pressure induced by a sound wave. Using microphone sensor provides a non-intrusive manner for sensing as these sensors do not need contact to measure the change in equipment condition [90]. However, in both AE and microphone sensing, these sensing methods are susceptible to noise from surrounding systems. One method to reduce the noise is to filter the condition monitoring data upon collection to remove interfering background noise elements [91]. Another method involves the use of adaptive noise cancelling using an auxiliary or reference signal to filter out the data [92]. Another potential method is the use of physical filtering to remove background noise through the design of the system [93]. The noise mitigation from these different

methods increases the likelihood of a correct diagnosis for the equipment.

2.4.4   Temperature

Temperature is a physical quantity used to express quantitatively different levels of hot and cold. In bearing condition monitoring, bearings are rated to operate under certain levels depending on the application. Detweiler [94] notes that bearing temperatures change based on the application. For example, electric motors typically operate between 140°F – 160°F (60°C – 71°C). In gear drives, bearing temperatures can range from 160°F - 180°F (71°C – 82°C). The change in temperature a bearing may experience comes from either a change in the lubrication conditions, resulting in further metal-on-metal contact. The bearing heating increases the rate of lubrication degradation and potentially early failure in the system.

Bearing temperature is monitored through the use of a sensor attached as close to the bearing as possible to acquire an accurate reading. These sensors are generally small and comprise of thermocouples [95] or thermistors [96] to measure the temperature change in the system. A thermocouple produces a temperature dependent voltage using the Seebeck effect [97] by using two dissimilar electrical conductors. Thermocouples exist in different forms based on the materials used to create the conductors [98]. Examples of the K, J, and E thermocouples are demonstrated in Figure 2.10. For the best measurement, the thermocouples should contact the bearing. In the case of electric motors, 15°F can be added to the temperature reading if it the sensor does not clearly contact the bearing [94]. Different thermocouple types may perform better depending on the application. Thermistors are resistors, where the resistance is dependent on the temperature. Thermistors are divided into Negative Temperature Coefficient (NTC) and Positive Temperature Coefficient (PTC) thermistors [99]. For NTC thermistors, resistance decreases as temperature rises. For PTC thermistors, resistance increases as temperature rises. These thermistors are placed in line with a circuit and used to measure the voltage drop over the thermistor. The voltage change corresponds then to the temperature rises.

Temperature is a common application in condition monitoring due to the direct correlation between equipment condition and temperature changes. Temperature condition monitoring defines set limits that inform maintenance operators in condition monitoring. Albers *et al.* [100] considers three different levels for motor bearing health, defined as "healthy", "warning", and "shutdown". As temperature increases, maintenance operators can take appropriate actions to stop failures from
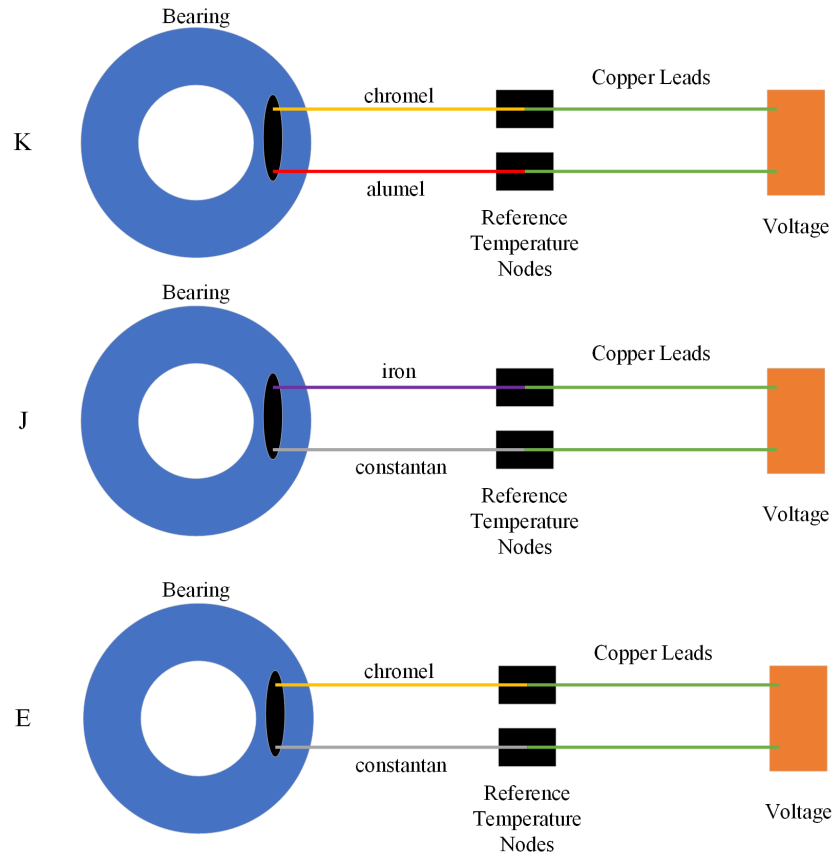
Figure 2.10: Examples for the K, J, and E type thermocouples

occurring. While temperature monitoring provides clearly defined limits for equipment conditions, it is noted that temperatures can rise from other elements in the operating environments [101]. For example, if the temperature sensor is exposed to the sun, that would naturally cause a temperature increase not related to the equipment condition. To mitigate this phenomena, temperature monitoring is paired with other sensor types, such as vibration [102] to increase fault diagnosis accuracy.

2.4.5   Ultrasonic Sensing

Ultrasonic sensing measures sound waves generated in the ultrasonic frequency range. Usually, the ultrasonic frequency range is defined as sound waves above 20 kHz [103]. Figure 2.11 places ultrasound in reference to other audio ranges. Acoustic ranges were discussed earlier in Section 2.4.3 using microphone and sound sensing and represent what is audible to humans. It should be noted

that with AE sensing, there is overlap in the frequency ranges for sensing between the two technologies. Vibration sensors can also measure ultrasonic frequencies if tunable to that sampling rate and frequency range. Ultrasounds and infrasound ranges [104] are defined as inaudible to humans. Infrasound is not necessarily useful in condition monitoring, since faults, especially bearings, occur above the 20 Hz range.

Figure 2.11: Audio range classifications for infrasound, acoustic, and ultrasound

Utrasonic sensing is usually conducted using either a ultrasound probe as done in Kim *et al.* [105] or transducer as done in Drinkwater *et al.* [106]. For Kim *et al.* [105], the ultrasound probe was an industrial solution that comprises an all in one system for both the sensing device and signal conditioner. The usable range for the probe was between 20 and 100 kHz for measurement. For the transducer solutions in Drinkwater *et al.*, the center frequency range was 200 MHz. However, additional components were needed such as a signal generator and a pulser-reciever to capture the ultrasound signals. In either situation, the data are passed back through a DAC prior to analysis with a computer.

The benefits of ultrasonic sensing primarily come from measuring lubrication deviations [107, 108] and early inception defects, such as cracks, as done with AE sensing [109]. The lubrication film for bearings is very fine on the scale of micrometers. If there is too little lubrication, the bearing rolling elements contact with the bearing raceway and have direct metal-on-metal contact. The contact increases the amount of friction and stress between the rolling elements and speeds up degradation. Zhang *et al.* [108] detected within 90% of the theoretical solution with loads greater than 1.5 kN. Another benefit for ultrasound sensing is the increased accuracy for bearing detection at slower speeds for bearings [105]. Vibration has a harder time detecting the impulses related to bearing defects, due to the time related to turning the shaft.

2.4.6   Data Acquisition: Synthesis

A literature review was conducted to determine the breakdown of condition monitoring over these five technologies. Section B contains the initial search terms used to conduct the liter-

ature review. A total of 10 search terms were used comprising the data acquisition method (*i.e.,* "Vibration") and either the term "Rolling Element Bearing" or "Ball Bearing". The search was conducted over papers from the last 5 years from the Compendex database. The papers were further paired down based on the search vocabulary referring to a physical system (*i.e., "magnetic bearing", "rotors", "aircraft engines"*). In addition, the top 10 journals were selected along with the five top publishers. Table 2.3 contains the breakdown of condition article with respect to the data acquisition method and application. Appendix B contains a further breakdown of the methodology.

Table 2.3:   Rolling element bearing condition monitoring data acquisition methods and application

Data Acquisition Method	References	Applications
Vibration	[110], [111], [112], [113], [114], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129], [130], [131], [132], [133], [134], [135], [136], [137], [138], [139], [140], [141], [142], [143], [144], [145], [146], [147], [148], [149], [150], [151], [152], [153], [154], [155], [156], [157], [158], [159], [160], [161], [162], [163], [164], [165], [166], [167], [168], [169], [170], [171], [172], [159], [173], [174], [175], [176], [177], [178], [179], [180], [181], [182], [183], [184], [185], [186], [187], [188], [189], [185], [190], [191], [192], [193], [194], [195], [196], [197]	*Bearing Test Rig:* [110], [112], [113], [114], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129], [130], [131], [132], [133], [135], [136], [137], [139], [140], [138], [141], [142], [143], [144], [198], [145], [146], [147], [148], [149], [150], [151], [199], [152], [153], [154], [155], [156], [157], [158], [160], [161], [162], [163], [164], [165], [166], [200], [167], [168], [169], [170], [171], [172], [159], [173], [174], [175], [176], [177], [178], [181], [182], [183], [184], [185], [187], [188], [189], [201], [202], [192], [193], [194], [195], [196], [197],

Vibration (continued)	[203], [204], [205], [206], [207], [208], [209], [210], [211], [212], [213], [214], [215], [216], [217], [218], [219], [220], [221], [222], [223], [224], [201], [225], [226], [227], [228], [229], [230], [231], [232], [233], [234], [235], [179], [179], [180], [190], [186], [236], [237], [198], [111], [238], [191], [239], [240], [202], [199], [241], [242], [243], [244], [245], [246], [247], [248], [200], [249], [250], [251], [252], [253],	*Bearing Test Rig (continued)*: [203], [204], [205], [206], [207], [208], *Datasets:* [209], [210], [211], [202], [126], [127], [111], [238], [212], [213], [214], [215], [145], [199], [159], [216], [217], [218], [219], [220], [221], [222], [223], [224], [201], [225], [190], [226], [227], [228], [229], [230], [231], [232], [233], CNC: [234], [235], [134], [179], [180], [190], [186], [254], [255] *Bearing Test Rig (Ceramic):* [236], [237], *Bearing Test Rig (Magnetic):* [198], *Simulation:* [111], [238], [191], *Gearbox:* [239], [240], [202], [199], [241], [242], [243], [244], [245], *Railway:* [246], [247], [248], [200], [254], [255], [249], *Wind Turbine:* [111], [250], [251], *Pump:* [252], [253]
Electrical Current	[256], [257], [258], [259]	*Bearing Test Rig:* [256], [257], [258], [259]
Acoustic Emission and Sound	[256], [260], [261], [110], [262], [136], [162], [252], [205], [208]	*Bearing Test Rig:* [262], [256], [261], [110], [136], [162], [205], [208], *Bearing Test Rig (Ceramic):* [260], *Pump:*[252]

Temperature	[260], [236], [128], [213], [136], [146], [169], [189], [230], [205], [263], [264], [207], [265], [266], [208], [267]	*Bearing Test Rig:* [136], [146], [169], [189], [205], [263], [264], [207], [265], [266], [208], [267], *Bearing Test Rig (ceramic):* [260], [236], [128], [213], [230]
Ultrasonic	[268]	*Wind Turbine:* [268]

Vibration is still the most prevalent condition monitoring technology available in research and industrial applications. Out of 170 articles reviewed from the initial search, 154 articles utilized vibration condition monitoring for bearing fault diagnosis. Temperature is the next largest category at 17 articles; it should be noted that many of the articles utilized both technologies. In the event that two or more technologies in the article, the reference was repeated for both data acquisition methods. A potential reason for the lack of ultrasonic could stem from the training required to utilize these technologies for condition monitoring. Kim *et al.* [105] notes that some tuning was required for the ultrasound probe. In addition, Drinkwater *et al.* [106], depending on the solution used, noted the level of precision and equipment needed to integrate their transducer in the bearing. It was surprising to see not as many papers related to current CM data acquisition, given the relatively un-intrusive method for sensing in motors. A potential reason could stem from the filtering out of sources from the search process as many of the journals were mechanical in nature; however, IEEE was one of the publishers used. The review was conducted to emphasize how each of these technologies are used in research and to inform the reader on the breakdown of technologies.

2.5 Data Preprocessing, Processing, and Analysis

The overlap between data processing and analysis is used interchangeably in CM research literature. If one were to continue with the CBM process laid out by Jardine *et al.* [71], data processing is the conversion of the data into features that determine the equipment health state. Data analysis comprises a "Maintenance decision support" system that determines the equipment health state based on the features. Jardine *et al.* describes some of these methods as time and frequency articles. In some research articles, the transformation process between time and frequency space is a data preprocessing technique [269, 270]. In other works, data processing encompasses the

entire process from data collection to the fault determination [271, 272]. To try and standardize the process, Figure 2.12 lays out the flow data from the data generation to the data analysis section.
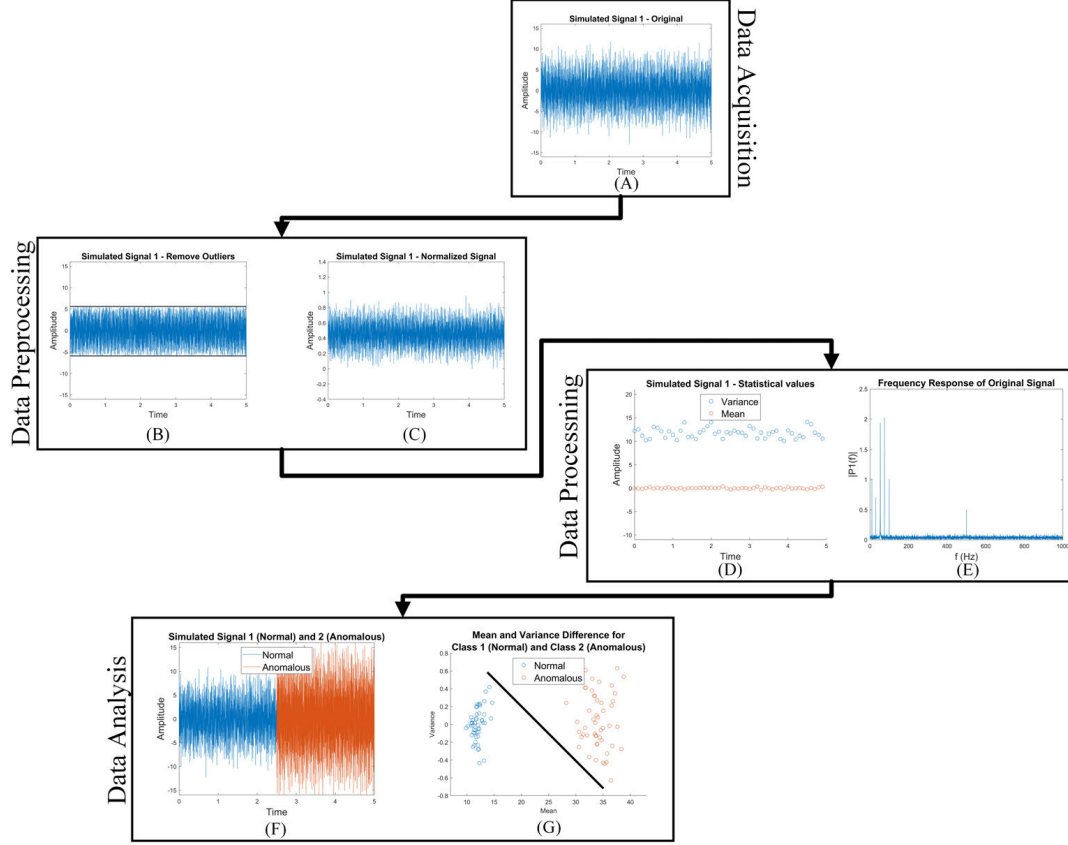
Figure 2.12: Layout of data preprocessing, processing, and analysis after data acquisition. (a) original collected signal, (b) removed outliers, (c) normalized data, (d) statistical values, (e) frequency values, (f) two separate healthy and anomalous signals, (g) identified clusters based on statistical features and identified hyperplane

This section standardizes these definitions to ensure this research work is clear about certain data types and their relationship to different parts of the condition monitoring system. Data preprocessing refers to any data preparation techniques before feature generation. As mentioned, this term changes between preprocessing and processing when considering data transformations (*i.e.*, time to frequency data). For this work, any data manipulation that does not result in a direct feature for the data analysis is considered a preprocessing technique. Data preprocessing could consider techniques such as data normalization or data outlier reduction. Data processing covers the conversion of data into features, also known as feature generation [273]. Feature generation is a term that comes from machine learning implementation and involves condensing data into usable metrics to differentiate

37

classes. Data processing uses time-based measures, such as the root mean square (RMS) and Kurtosis, or frequency peaks, to reduce the data for analysis. Data analysis involves using data techniques to determine the correct data class associated with the original data. Based on the selected category, information becomes available for effective decision-making. Data analysis techniques include machine learning applications and physics-based modeling. The subsections comprising the remainder of the section discuss and cover the areas of data preprocessing, processing, and analysis.

2.5.1   Preprocessing

As mentioned earlier, data preprocessing deals with any data manipulation that does not result in a feature. Data that comes directly from the data source generally has errors. A sensor could fail unexpectedly, an interruption could occur in the data transmission, or human error may have incorrectly configured the equipment before collection. Data preprocessing is needed in each scenario to mitigate and correct potential errors that may otherwise inhibit accurate classification. Famili *et al.* [274] notes these scenarios in their review and discussion for data preprocessing, specifically in the case of "too much" data, "too little" data, or "fractured" data. A data scenario with too much data may have additional noise or irrelevant points mixed in with the expected data. A system with too little data may have missing values from the data. A fractured data set may have incompatible data types mixing from multiple sources at different levels. Hence, measures are needed to account for these data inconsistencies.

Alsadi *et al.* [275] describes several different techniques, such as manually filling in the data through either a global or probable constant based on the data source. Another technique is binning to reduce the effect of minor data observations. Ramirez-Gallego *et al.* [276] considered several different data reduction scenarios to reduce the data amount potentially affecting the overall analysis for online processing and analysis. They primarily investigated using filters and wrappers to reduce the data amount during their work. Alexandropoulos *et al.* [277] highlighted the use of noise and outlier detectors to limit their impact on the generation of features. They also brought up the idea of feature selection and methods to mitigate missing feature values. While some may argue that this means data preprocessing and data processing should remain interchangeable, feature selection, or feature extraction as it is sometimes known, is encompassed in the data processing section.

Many different preprocessing techniques are used in condition monitoring, especially bearings, due to the wide degree of variance accompanying manufacturing data. A manufacturing en-

vironment is considered harsh, with additional chances of accumulating damage depending on the process and industry. Hence, data sources experience similar harshness levels related to their equipment and can accumulate other noise. One immediate method for data cleaning and reduction is filtering. Data filtering uses digital or physical methods to remove specific data from the collected signal. Han *et al.* [278] highlights several for electrical equipment (motors and transformers), such as ban pass filtering, low pass filtering and high pass filtering. Figure 2.13 considers three different filtering methods to eliminate data outside of a specific frequency. A low pass filter allows only signals below the cutoff frequency through the filter, whereas a high pass filter will enable signals above a particular frequency for analysis. The band-pass filter considers only signals between a lower and upper frequency. By applying the correct filter, certain undesirable data bands are removed from the feature generation. Filters can be used digitally or physically, depending on the application. A physical filter would involve modifying a physical component to reduce the amount of noise from surrounding objects. Skoglind *et al.* constructed a physical filter applied to a microphone to detect quality checks in electrical connections. Microphones can also be built with directionality to eliminate surrounding ambient noise that may corrupt a signal [279]. Pointing the microphone at the equipment ensures that the microphone collects only the expected signals from the equipment and reduces the effect of the noise from outside sources.
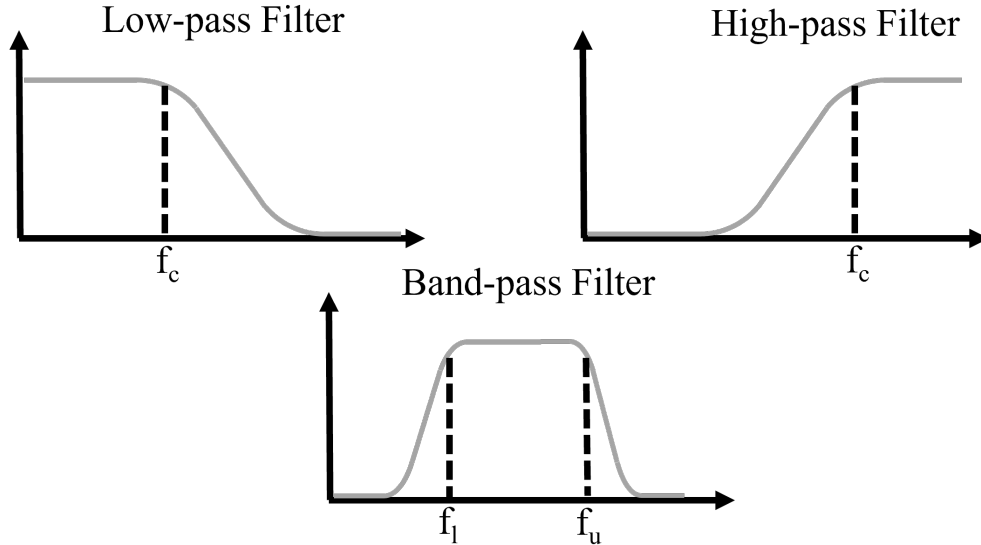
Figure 2.13: Different filtering methods: Low-pass filter, High-pass filter, and Band-pass filter

The cutoff frequencies chosen for the above filtering methods are generally based on known

expected frequencies or experimentation. In practice for bearings, four different critical frequencies require monitoring. These defect frequencies are associated with the other bearing parts that are discussed later in this section. In addition to these frequencies, speed frequencies are captured depending on how fast the bearing is rotating. Depending on the application, other frequencies, such as the motor slip, speed, and rotor bar frequency, could be captured. Filters generally isolate and remove potential extraneous noise around the expected bearing frequencies.

In addition to data filtering, data transformations help visualize and identify signals of interest that occur at specific instances. For condition monitoring, time series data are often transformed into frequency space data because specific frequencies increase in amplitude as equipment degrades. For bearings, the features generated from the frequency space could correspond to increases in defects along the bearing raceway or an increase in contamination in the bearing lubrication. Frequency transformations in condition monitoring used are the Fast Fourier Transform (FFT), the Hilbert-Huang Transform (HHT), and Wavelet Transforms (WT).

The Fast Fourier Transform (FFT) is a method for computing the discrete Fourier transform of time series data [280]. From the transformation, it is possible to view the different amplitudes associated with each frequency value between zero and half of the sampling rate. For example, if the sampling rate is 1000 Hz or 1 kS/s, the maximum possible frequency to analyze is 500 Hz. In addition to the maximum measured frequency, the spacing between frequency points is called the frequency resolution. The frequency resolution is determined based on the maximum sampling rate divided by the number of available samples [281]. A higher resolution FFT allows for easier viewing of specific frequencies but requires more significant sample amounts. FFTs are used in condition monitoring due to their efficient computational implementation. After the transformation, different signal parts are analyzed to determine the equipment health state [282]. However, a drawback to the FFT is that it also does not handle changes to the operating conditions with equipment. With other techniques, such as the HHT and WT, the time series information is still captured in the signal, allowing for better resolution as the operating conditions change over time.

$$Y(K) = \sum_{j=1}^{n} X(j) W_n^{(j-1)(k-1)} \tag{2.6}$$

Hilbert Huang Transform (HHT) is another method for extracting noisy signals using empirical mode decomposition (EMD) to extract the noise from the intrinsic mode functions (IMFs)

[283]. EMD is a method to decompose a signal into separate components related to different system parts. These individual components are called intrinsic mode functions. IMFs are characterized by two properties: (1) the number of extrema and zero crossing points must be equal or differ by one, and (2) the mean value of local minima and maxima must be zero [284]. The decomposition of the signal into the IMFs allows for the analysis of systems with non-stationary conditions. Non-stationary conditions mean operating conditions that change concerning time or equipment state. Since bearing condition monitoring requires knowledge of the equipment characteristics, such as load and speed, the HHT helps determine these conditions and tracks their changes in the time and frequency domain.

Empirical mode decomposition involves sifting to extract the usable IMFs from the time-series equation. The sifting process aims to eliminate the extraneous and outlier points and have the data appear more symmetric. Hence, the data is sifted until it reaches the above IMF condition. The general procedure follows as shown below in Equation 2.7. $X(t)$ represents the original raw data, whereas m_1 represents the mean of the upper and lower envelope. The mean signal is subtracted from the algorithm to form the first component of the sift, h_1. h_1 is then used as the next part of the sift. The algorithm is used successively until it satisfies the conditions for the IMF. Consequently, h_{1k} replaces $X(t)$, and the process begins again until no more IMFs are extracted.

$$X(t) - m_1 = h_1$$
$$h_1 - m_{11} = h_{11}$$
$$h_{1(k-1)} - m_{1k} = h_{1k} \tag{2.7}$$
$$IMF_17 = h_{1k}$$

The frequency transform then employed is the Hilbert Transform (HT), where the main equation is provided as shown in Equation 2.8. The Hilbert transform is used due to the interest in obtaining the instantaneous frequency, $Y(t)$. P represents the Cauchy principal value and $X(t)$ represents the raw time series values. The capture of the instantaneous frequency using the HT allows for the matching of frequencies to their associated time series expectations.

$$Y(t) = \frac{1}{\pi} P \int_{-\inf}^{\inf} \frac{X(t')}{t - t'} dt' \tag{2.8}$$

The final discussed method are Wavelet Transforms. A wavelet is an oscillation mimicking a wave that begins at zero and increases or decreases multiple times [285]. These wavelets are passed over data numerous times to capture frequency information while retaining the time data. The most common application for condition monitoring literature is the Continuous Wavelet Transforms (CWT). A continuous wavelet transform considers the representation of a signal by constantly varying the translation and scale parameters of a wavelet to represent a time-based signal. Grossman and Morlet first implemented the process of continuous wavelet transforms [286]. They subsequently created the Morlet Wavelet is characterized by a complex exponential and Gaussian window. While wavelets are typically employed, generally, there implementation requires extensive experimentation to select the appropriate wavelet for each scenario. In this essence, wavelets are considered more of a future item rather than employed with this work.

Data transformations for transferring data from the time to frequency space are common applications alongside data filtering methods. Other methods to reduce the error amount in a signal for condition monitoring are similar to handling data scenarios as described by Famili *et al.* [274]. Bangalore *et al.* described the use of different "filters" to account for missing data and add a cluster probability to the data. By adding these additional filters for the analysis, they were able to ensure continuity of data prior to analysis in the model. Yang *et al.* [287] and Ompusunggu *et al.* [288] used data binning to reduce the data dimensionality prior to failure analysis. The subsequent data binning allowed for the feature generation in a streamlined manner for online condition monitoring. Wu *et al.* [289] considered data normalization of AE signals to allow for direct comparison related to other sensors and overcome the nonlinear response problems. Yang *et al.* [290] considers the use of data normalization to scale the data prior to use in their analysis. They suggest that some of their objective functions for condition monitoring would not work without scaling the data.

For condition monitoring, there are different methods to ensure that the data are presented correctly before analysis. The proper data presentation each time leads to a smoother transition from the data processing to the data analysis stage. The preprocessing stage also reduces the possibility of "garbage" data entering the processing and analysis stage. Generally, the rule of thumb for data processing and analysis is "Garbage In equals Garbage out". The phrase means that if the data are not helpful, the information generated is equally useless for any equipment analysis.

2.5.2 Processing

Data processing encompasses the techniques for creating features and analysis that indicate possible equipment damage. Equipment degradation, barring a catastrophic event (*i.e.,* human error or inadequate equipment design or function), does not immediately occur without some indication in the data to signal the possible degradation to the failure point. These features are usually processed in either the time or frequency domain. The time domain represents data collected concerning time, whereas the frequency domain contains data concerning frequency. Time domain features generally consider statistical values to differentiate between healthy and failed equipment. Dyer *et al.* [291] tracked the change of RMS and Kurtosis over time to match the degradation of equipment. Mechefske [292] described that the overall data distribution changed as equipment degraded, as shown in Figure 2.14. The data distribution changes will change depending on the different defect associated with the bearing. Analyzing raw vibration signals makes it possible to pick out faults based on where bearing impulses occur [293].
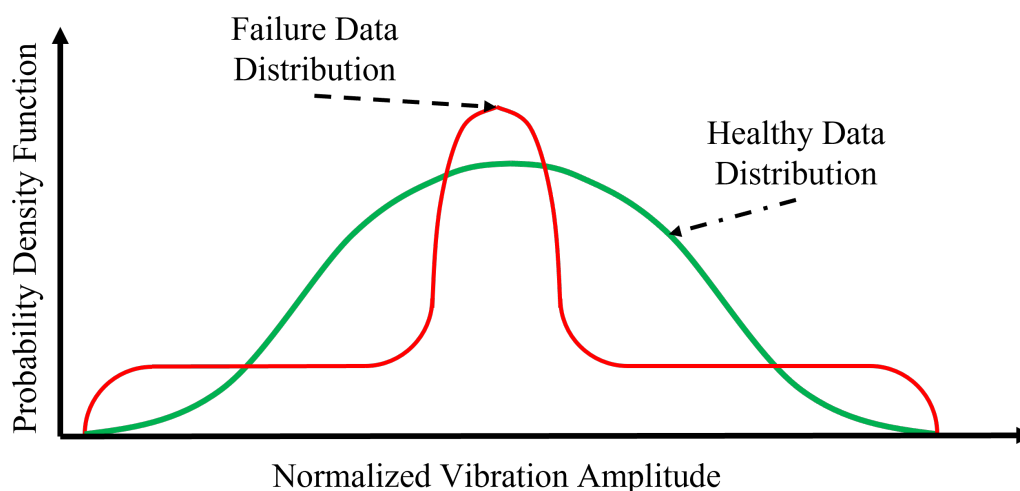
Figure 2.14: Representative changing data distribution based on the equipment degradation

The change in the kurtosis could stem from additional outliers in the bearing data. Kurtosis increases from the other data that comes from the outliers. That makes the data distribution "heavy-tailed". The additional outliers could stem from the increased defect events causing vibration spikes. As these events grow, the damage accumulates on the bearing surface. The increased damage on the bearing surface registers additional spikes as the bearing rolling element goes over the bearing

surface.

The changes in data distributions are usually described by statistical parameters, such as the RMS, Variance, Kurtosis, and Skewness of the data [294]. The RMS measures the overall level of a discrete signal and can measure the vibration power. This feature differs from the signal mean, which only works for rectified signals. For standard time series data, the signal remains close to zero. The variance of the signal tracks the dispersion from the average expected value. The Kurtosis and skewness of the data describe the change in normal distributions. As bearings degrade, the data distribution also changes from a normal state. Additional measures consider the peak acceleration and crest factor. The peak acceleration reports the maximum possible value from the bearing. The crest factor is the ratio of the peak acceleration to the RMS. The deviations and trends associated with the crest factor are considered unreliable, though. Equations 2.9 through 2.14 shows the equations for calculating the RMS (2.9), mean (2.10), variance (2.11), Kurtosis (2.12), skewness (2.13), and crest factor (2.14). x_i refers to the individual samples in a group of samples x. N is the number of samples, and $\bar{x}$ is the mean.

$$RMS(x) = \sqrt{\frac{\sum x_i^2}{N}} \tag{2.9}$$

$$\bar{x} = \frac{1}{N} \sum_{N}^{i=1} x_i \tag{2.10}$$

$$VAR(x) = \frac{\sum (x_i - \bar{x})^2}{N - 1} \tag{2.11}$$

$$Kurtosis = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4}{((N - 1) \sum_{i=1}^{N} (x_i - \bar{x})^2)} \tag{2.12}$$

$$Skewness = \frac{1}{(N - 1)} \sum_{i=1}^{N} (x_i - \bar{x})^3 \tag{2.13}$$

$$Crest = \frac{PeakAcceleration(PA)}{RMS(x)} \tag{2.14}$$

Helmi *et al.* [293], Kundu *et al.* [230], Ren *et al.* [295], and Caesarendra *et al.* [296] considers similar time features in their different analysis techniques for bearing condition monitoring

and remaining useful life prediction. Caesarendra *et al.* and Helmi *et al.* provide additional terms as well. Caesarendra *et al.* uses a shape factor and characterizes it as a non-dimensional feature, similar to the crest factor for determining bearing faults. In addition, they used the entropy of the vibration signal to determine the uncertainty of random vibration signals. Equation 2.15 and 2.16 are the shape factor and entropy equation as called out by Caesarendra *et al.*. The entropy equation is characterized by the probabilities p based on a vibration signal x_i. In addition to these equations, Helmi *et al.* considered the standard deviation, square mean root, margin factor, impulse factor, mean absolute, standard deviation absolute, skewness absolute, and Kurtosis absolute features in bearing condition monitoring. Equations 2.17 and 2.18 are the standard deviation (SD) and square mean root (SMR), respectively. These parameters are variations of the variance and RMS. Equations 2.19 and 2.20 are the margin factor and impulse factor, respectively. These parameters are similar to the shape and crest factor and are considered non-dimensional features. Equations 2.10 through 2.24 describe the absolute value variations from earlier described parameters. Helmi *et al.* does not explain the impact these features have in determining bearing fault over the conventional features as listed for Equations 2.9 through 2.13 and Equation 2.17.

Nayana *et al.* [297] did provide a comparison for their proposed features: the mean absolute (MA), the simple sign integral (SSI), the waveform length (WL), the Wilison Amplitude (WAMP), the zero crossing (ZC), and the slope sign change (SSC). Equations 2.25 through 2.32 are the equations for the SSI (2.25), the WL (2.26), the WAMP (2.27) and the corresponding conditional (2.28), the ZC (2.29) and the corresponding conditional (2.30), and the SSC (2.31) and the corresponding conditional (2.32). The SSI is another variation of the MA to describe the signal amplitude over a vibration sample. The WL and WAMP describe the frequency of variation between individual samples. The WAMP provides the variation provided within a certain noise level, defined as ϵ. The ZC and SSC calculate the frequency signals corresponding to specific vibration samples. Nayana *et al.* compared the accuracy using these features against the conventional features in a NB and SVM classifier. They were able to find similar accuracy levels between either feature group.

$$SF = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}}{\frac{1}{N}\sum_{i=1}^{N}\|x_i\|} \tag{2.15}$$

$$e(p) = -\sum_{i=1}^{n}p(x_i)log_2p(x_i) \tag{2.16}$$

$$Standard\ Deviation\ (SD) = \sqrt{(VAR(x))} \tag{2.17}$$

$$Square\ Mean\ Root\ (SMR) = \sqrt{\frac{\sum_{n=1}^{N}(x_i)^2}{N}} \tag{2.18}$$

$$Margin\ Factor\ (MF) = \frac{PA}{SMR} \tag{2.19}$$

$$Impulse\ Factor\ (IF) = \frac{RMS}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{2.20}$$

$$Mean\ Absolute\ (MA) = \frac{\sum_{i=1}^{N}|x_i|}{N} \tag{2.21}$$

$$Standard\ Deviation\ Absolute (SDA) = \sqrt{\frac{\sum(|x_i|-\bar{x})^2}{N}} \tag{2.22}$$

$$Skewness\ Absolute\ (SA) = \frac{1}{(N-1)SD^3}\sum_{i=1}^{N}(|x_i|-\bar{x})^3 \tag{2.23}$$

$$Kurtosis\ Absolute\ (KA) = \frac{\sum_{i=1}^{N}(|x_i|-\bar{x})^4}{(N-1)(\sum_{i=1}^{N}(x_i-\bar{x})^2)} \tag{2.24}$$

$$SSI = \sum_{i=1}^{N}|x_i|^2 \tag{2.25}$$

$$WL = \sum_{i=1}^{N}|x_i - x_{i-1}| \tag{2.26}$$

$$WAMP = \sum_{i=1}^{N}f(|x_i - x_{i+1}|) \tag{2.27}$$

$$f(x) = \begin{cases} 1 & x \geq \epsilon \\ 0 & otherwise \end{cases} \tag{2.28}$$

$$ZC = (x_i > 0 \text{ \&\& } x_{i+1} < 0)||(x_i < 0 \text{ \&\& } x_{i+1} > 0) \tag{2.29}$$

$$|x_i - x_{i+1}| \geq \epsilon \tag{2.30}$$

$$SSC = (x_i > x_{i-1} \text{ \&\& } x_i > xi + 1)||(x_i < x_{i-1} \text{ \&\& } x_i < x_{i+1}) \tag{2.31}$$

$$|x_i - x_{i-1}| \geq \epsilon \tag{2.32}$$

In addition to the time features described, frequency features track degradation at specific frequencies based on changing amplitude and energy content across a frequency range. These frequencies correlate to specific physical components that exhibit degradation. Buhl *et al.* [298] used frequency analysis to identify the difference between clean and dirty lubrication for bearing conditions. NASA contracted researchers to investigate bearing failure utilizing frequency analysis for different bearing faults [299]. Observing these frequencies for bearing condition monitoring has formulated impulse theory found in physics-based modeling for bearing analysis [300, 301, 302]. Frequency spectra have also been converted to images for deep learning algorithms to identify bearing fault diagnosis [303]. Depending on the different faults, the images exhibit additional frequency information as samples are collected. There are commonly accepted failure frequencies related to the bearing geometry and speed for bearings. Equations 2.33 through 2.36 describe the calculation for the Ball Pass Frequency Inner (BPFI) (2.33), the Ball Pass Frequency Outer (BPFO) (2.34), the Fundamental Train Frequency (FTF) (2.35), and the Ball Spin Frequency (BSF) (2.36). The BPFI and BPFO correspond to defects along the raceway of the inner and outer race, respectively. The FTF and BSF correspond to defects on the cage and rolling elements, respectively. N refers to the number of balls, and B is the ball diameter in mm. P is the bearing pitch diameter in mm, calculated by using the mean of the bearing bore diameter (shaft diameter) and outer diameter (bearing outer ring diameter). ϕ is the contact angle for the rolling elements to the bearing raceway. Ω is the bearing speed measured in Hz. The bearing frequencies are expressed in Hz if they include the bearing speed. If the bearing speed is not included, a dimensionless value termed undulations per revolution (UPR) describes the bearing frequency. Generally, as bearing defects appear on certain

elements, these frequencies increase in amplitude. Additionally, their harmonic frequency values may increase in amplitude as bearings degrade.

$$BPFI = \frac{N}{2}\Omega(1 + \frac{B}{P}\cos\phi) \tag{2.33}$$

$$BPFO = \frac{N}{2}\Omega(1 - \frac{B}{P}\cos\phi) \tag{2.34}$$

$$FTF = \frac{\Omega}{2}(1 - \frac{B}{P}\cos\phi) \tag{2.35}$$

$$BSF = \frac{P\Omega}{2B}(1 - (\frac{B}{P})^2(\cos\phi)^2) \tag{2.36}$$

For industrial applications, vibration and sound measurement techniques are tuned to capture these vibration signals. The maximum measured frequency must be below 1/2 the sampling rate, which is the folding or Nyquist frequency. The effects of aliasing are reduced by meeting this criterion. Typically, for most bearing signals, ensuring the sampling rate is at least ten times the last critical frequency is good practice. For example, if a bearing has a maximum fault frequency of 100 Hz, the minimum sampling rate should be 1000 Hz. Mechefske *et al.* [292] detailed the use of expected bandwidth limits for frequency-based data. The method sets cutoff frequencies below and above the defect frequency at specific intervals. The equipment's remaining useful life decreases as deviations occur in these bandwidths. These frequencies are also captured in certain band limits designed by the person performing the analysis or the technique used [304]. However, these bands can depend on the method used to isolate the bearing frequency, as mentioned in Section 2.5.2. Non-stationary techniques are generally the preferred method as they reduce the possibility of low visualization for the bearing fault frequencies, especially in the machine and deep learning methods [305]. For these data-driven processes, hidden fault frequencies may not bring the expected results when considering potential noise in the bearing signal. Other band limits for frequency analysis are based on the bearing signal's intrinsic mode functions [306, 192]. Kurtosis can be calculated from the extracted frequencies for the IMF to determine the fault type and severity.

2.5.3 Analysis

The two primary methods used for bearing analysis are either data-driven methods, such as machine and deep learning, or physics-based methods, such as mass-spring-damper (MSD) systems. Table 2.4 contains review articles for bearing condition monitoring and fault diagnosis, comprising physics-based methods and data-driven methodologies. Data-driven modeling operates on training a model on a set of sample data to predict the outcome of any new data. The learning conducted from the data-driven model is purely based on the characteristics of the data patterns. Physics-based modeling is based on the physical relationships of the system that the data describes and represents. Typically, these models are more computationally intense; however, their predictions are related directly to the system's physical components versus indirectly measured data. Between these two modeling techniques, hybrid methods combine techniques from both domains.

2.5.3.1 Data-Driven Modeling

Data-driven modeling primarily utilizes machine learning models to predict and determine certain outcomes based on the train data. As mentioned earlier, the models learn entirely based on an input of sample data that trains and configures the model to predict certain outcomes. There are three learning approaches for data-driven modeling: Supervised, Unsupervised, and Semi-Supervised learning. Supervised learning algorithms create a function that predicts an outcome based on feature inputs and is trained using input-output pairs [318]. Essentially, the input-output pairs used for training are termed "labeled training data". The ideal algorithm from supervised learning can predict "unseen" or new data used in target analysis. For condition monitoring, support vector machines (SVM) [138], Naïve Bayes [319], k-Nearest Neighbors (kNN) [320], Decision Trees [320], Random Forest [320], and Artificial Neural Networks (ANN) [321] use supervised learning for training their respective approaches. The training size for constructing the algorithm influences the algorithm's performance. Large amounts of training data are required to ensure model accuracy and for the analysis to predict the correct outcome [322]. Another consideration is the trade-off between bias and variance in supervised learning [323]. A biased algorithm continuously incorrectly predicts a certain class; a high variance algorithm changes its prediction based on the selection of training data. Mitigating ways for the amount of bias and variance in a machine learning or data-driven model is using different hyperparameters to tune the model output and reduce potential errors.

Table 2.4: Review articles for bearing data analysis

Reference	Journal	Models and Methods
Cerrada et al. [307]	Mechanical Systems and Signal Processing	Artificial Neural Network, Support Vector machines, Hidden Markov Modeling, Linear Discriminant Analysis, Fuzzy Logic
El-Thalji et al. [308]	Mechanical Systems and Signal Processing	Artificial Neural Network, Fuzzy Logic, Support Vector Machines, Principal Component Analysis, Paris's Law, Forman Law, Fatigue Spall Model, Self-Organizing Map, Recurrent Neural Network, Deep Belief Network, Expert Systems
Lei et al. [309]	Mechanical Systems and Signal Processing	Artificial Neural Network, Expert Systems, Support Vector Machines, k-Nearest Neighbors, Probabilistic Gaussian Models, Decision Trees, Autoencoders, Deep Belief Network, Convolutional Neural Network
Liu et al. [310]	Mechanical Systems and Signal Processing	k-Nearest Neighbor, Naïve Bayes, Support Vector Machine, Artificial Neural Networks, Convolutional Neural Networks, Autoencoder, Restricted Boltzmann Machines, Deep Belief Networks
Hoang et al. [311]	Neurocomputing	Autoencoder, Restricted Boltzmann Machine, Convolutional Neural Network
Lei et al. [312]	Measurement	Finite Element, Mesh Stiffness, Analytical models, Support Vector Machines, k-Nearest Neighbors, Ordinal Ranking, Hidden Markov Models, Bayesian Networks
Hamadache et al. [313]	JMST Advances	Linear Discriminant Analysis, Support Vector Machine, k-Nearest Neighbor, Extreme Learning Machine, Artificial Neural Networks, Convolutional Neural Network, Recurrent Neural Network, Restricted Boltzmann Machine, Autoencoder
Peng et al. [314]	International Journal of Advanced Manufacturing Technology	First Principle Modeling, Parameter Estimation, Fuzzy Logic, Expert systems, Artificial Neural Networks, Bayesian Modeling, State Space Modeling, Hazard Rate Modeling, Gray Model
Zhang et al. [315]	IEEE Access	Artificial Neural Networks, Principal Component Analysis, k-Nearest Neighbor, Support Vector Machine, Linear Discriminant Analysis, Bayesian Networks, Ensemble Learning, Convolutional Neural Networks, Autoencoders, Deep Belief Neural Networks, Generative adversarial Network, Recurrent Neural Networks
Peng et al. [316]	Algorithms	k-Nearest Neighbors, Artificial Neural Network, Support Vector Machines, Random Forest, Convolutional Neural Network, Autoencoder, Deep Belief Network, Recursive Neural Network
Alshorman et al. [317]	Shock and Vibration	Bayesian Networks, Supprot Vector Machines, Artificial Neural Network, k-Nearest Neighbor, Neuro Fuzzy, Deep Neural Network

Unsupervised learning is the opposite of supervised learning. Unsupervised learning trains the model based on the self organization of the training data by assigning certain probabilities to different models [324]. Unsupervised learning algorithms learn based on mimicry of the data. The model then corrects it self based on the error in the system to then update weights and biases in the data. Different unsupervised learning models exhibit two different methods, clustering methods and

anomaly detection methods. Clustering methods seek to predict outcomes on input features based on different data groupings [325]. Examples of these methods involve k-Means, Hierarchical clustering, linkage clustering, and Fuzzy c-means. Anomaly detection consider the detection of outliers based on rare events [326]. The assumption of these different methods considers that outliers are few and have different attributes. Unsupervised learning methods using anomaly detection include the Isolation Forest [326] and Local Outlier Factor [327]. Note, it is possible to train algorithms that are designed for supervised learning with unsupervised learning methods [328, 329]. The difference is dependent based on the algorithm hyperparameters and structure. For the unsupervised learning approach to work, enough separability is needed between classes to predict the expected outcome.

Semi-supervised learning considers a combination of training based on a small amount of labeled training data and a large mount of unsupervised samples. Li *et al.* [330] considers two general issues related to semi-supervised learning in machine learning applications: Data Quality and Model Uncertainty. Data quality considers how relatable the labeled training data are to the unlabeled training data. Subsequently, another issue is the prediction of new testing data, if they are not represented in the training case. Gomes *et al.* [331] notes that scenarios of weak and inconsistently provided labeled training data hampers the algorithm predictability. Model uncertainty is caused by different trials and different training data selections. One particular problem is pseudo labels, which can increase performance degradation. Despite the issues, semis-supervised learning is a common method in condition monitoring due to the low amount of labeled training data for the algorithms. Bull *et al.* [332] used semi-supervised learning with a Gaussian Mixture Model to reduce classification accuracy. Yuan *et al.* [333] used manifold regularization to determine different classes for bearing defects using semi-supervised learning. Essentially, these models are trained using the small samples of labeled data intially and then strengthed using the unlabeled data. Figure 2.15 considers the three different approaches with respect to the training data used. If the training data are labeled, then a supervised learning approach is used. If the training data are unlabeled, then an unsupervised learning approach is used. If there is a mixture of labeled and unlabeled data, then a semi-supervised learning approach is used. Based on the training data and application, a different machine learning or data-driven model can be chosen to predict the model case.
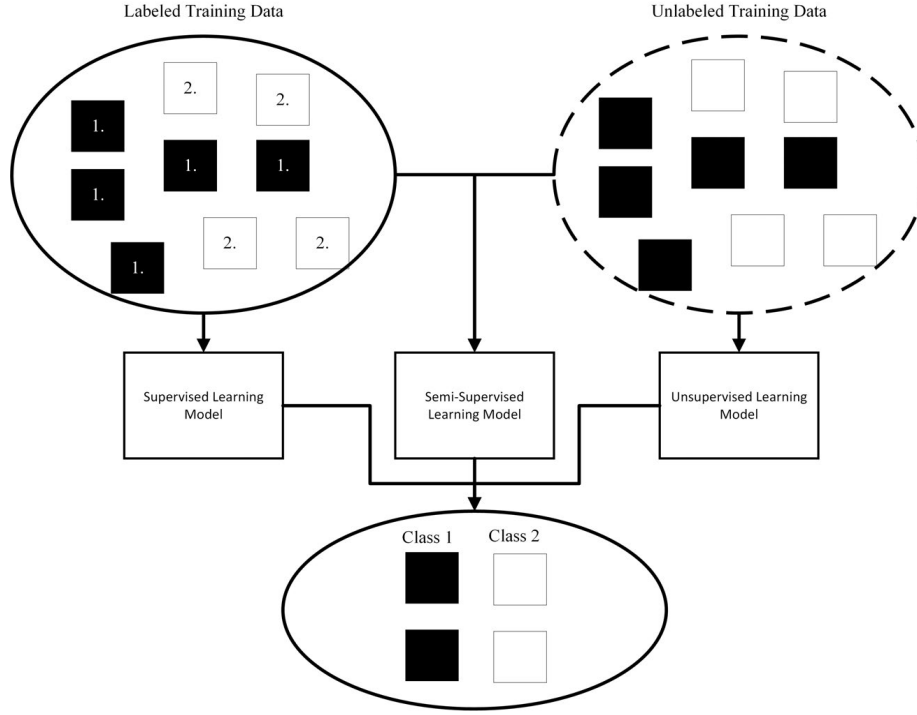
Figure 2.15: Learning approaches organized based on the training data used

2.5.3.2 Physics-Based Modeling

AS the name suggests, bearing physics models are rooted in the physics regarding the bearing system. The common physic principles for bearings are related to their design life as called out in ISO 281 [334]. The life modeling relations from ISO are determined based on the load and is commonly referred to as L10 life. L10 is the basic life rating for a bearing given in units of 10_6 rotations per hour. Equation 2.37 returns L10 life as referenced by the previous units. C is the dynamic load rating, and P is the dynamic application load. p is the exponential that is a data-driven parameter based on the bearing designation. Equation 2.38 considers the input of speed to provide an application specific life rating, where n is the number of rotations per minute. In other variations, an α is added for increased accuracy by including lubrication and contamintion information.

$$L_{10} = (\frac{C}{P})^p \tag{2.37}$$

$$L_{10h} = \frac{10^6}{60n}(\frac{C}{P})^p \tag{2.38}$$

These equations are derived based on data-driven experimentation for bearings. However, they form the basis for bearing life equations. For fault diagnosis methods, bearing impulses are used to simulate and model vibration responses under different operating conditions. Mcfadden *et al.* [302] is usually credited with the first impulse method for modeling bearing vibration as defects begin. Tandon *et al.* [335] used Lagrange's equations of motion to derive vibration response and frequencies based on the bearing rotation. The formulation of mechanics is based on stationary action principles as is commonly expressed as a variation of Equation 2.39. L represents a Lagrangian derived based on the change in kinetic and potential energy. x denotes the position particles that make up the Lagrangian and $\dot{x}$ denotes the velocity terms of the position particles. f_i denotes the external forces that may act on the system with respect to the position variables. Patel *et al.* used the Runge-Kutta equations to derive their equations of motions and to predict the bearing life relative to the different inner and outer defects. The Runge-Kutta equations are explicit and implicit methods to approximated nonlinear equations. Sassi *et al.* [336] created a numerical model to predict damaged bearing vibrations relative to the expected bearing vibration response using the finite element method (FEM). FEM is the process for approximating physical models using mathematical relationships. Liu *et al.* [337] constructed a two degree-of-freedom (DOF) model using Hertzian contact of the ball and the bearing defect raceway. Hertzian contact is used to describe the stress at the point of contact between two elastic bodies [338]. For ball bearings, the surface is modeled as a point; for roller bearings, it is modeled as a point. Sawalhi [339] considers a five degree-of-freedom (DOF) model for determining vibration response in the x and y direction for the bearing. The DOF of a bearing model are determined based on the number of input position points. The system DOF can be based on the bearing housing, shaft, the inner ring, the outer ring, or the rolling elements. Equation 2.40 provides the general representation for bearing equation motion that is derived based on the Lagrangian or Runge-Kutta methods and is the same as Equation 2.5.

$$\frac{\partial L}{\partial x_k} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x_k}} + \sum_{i=1}^{C}(\lambda_i)\frac{\partial f_i}{\partial x_k} = 0 \tag{2.39}$$

$$m\ddot{x} + b\dot{x} + kx = F \tag{2.40}$$

53

The previous methods described could be termed "motion models". They are formulated based on the equations of motions derived through different analytical methods. Outside of the equation of motions, laws have been utilized from fracture mechanics to describe defect propagation in bearings. One example is the use of the Paris Crack Growth (PCG) model, also nown as Paris Law, to determine the increase of bearing defect cracks over time [340, 341, 342]. The Paris crack law is formulated as shown in Equation 2.41. a is the crack length, where da/dN is the crack growth per loading cycle. C and m are coefficients determined as a resultant of the application in a particular environment. ΔK is determined based on the maximum and minimum stress in the model. Due to the constraints given by C, K, and M, the relationship serves as the basis for state models [341]. Qian *et al.* [341] combined PCG with a Hidden Markov model (HMM) to determine the remaining useful life over time for bearings. A number of other crack growth equations exist with respect to the Forman equation [343] and Elber equation [344]. The vast majority of bearing physics-based literature is represented through the Paris crack growth model.

$$\frac{da}{dN} = C(\Delta K)^m \tag{2.41}$$

2.5.3.3 Diagnostic vs. Prognostic Method

Bearing analysis is split into diagnostic and prognostic methods. Bearing diagnostics identifies a condition under certain scenarios, such as operating conditions or applications. The condition could be the different failure modes from the mechanical system or an anomaly from a baseline scenario. Essentially, the diagnostic is to detect when something has deviated from a normal state. The bearing prognostics are focused on determining the time until bearing failure. The bearing prognosis is based on the changes to the diagnosis over time. Vogl *et al.* [22] splits these in their analysis of current prognostics and health management (PHM) to consider the challenges in both systems. Ostensibly, these two methods should go hand in hand, as it happens in the medical field [345]. Diagnosis is viewed as the first step towards decision-making and generally is followed by a prognosis of treatment for the body to return to a natural state.

Data-driven models and physics-based models have been used in both mechanical system diagnosis and prognosis. For bearing diagnosis, classifier tasks generally are able to pick out the differences between certain failure criteria. Zhang *et al.* [315] primarily listed machine learning tasks, such as artificial neural network (ANN) and k-Nearest Neighbors (kNN), to pick out bearing faults

on the different rolling elements. Hybrid-based methods, the combination of physics and data-driven methods, are also used to implement bearing diagnosis. Sadoughi *et al.* [346] incorporated bearing fault frequency information into Convolutional Neural Network (CNN) to simulate and compare the spectral analysis between convolutions in the neural network architecture. Combining the physics information with data-driven modeling reduces the inherent black-box nature that accompanies machine learning and data-driven models. Physics-based methods are usually considered corroboration and identification for different bearing diagnostic models. Mcfadden's models [302, 301] are generally the basis for any physics verification for any fault generate bearing data using the impulse of rolling element striking the bearing raceway.

For bearing prognostic models, state-space modeling, forecasting, or regression analysis are used to determine trends as bearings degrade. One example is the use of HMM to estimate bearing state and remaining useful life [347]. The HMM uses the inputs as different states to determine potential outcomes. From Tobon-Mejia *et al.* [347] and Soave *et al.* [348], the HMM process takes in certain vibration levels and from there determine different potential outcomes related to remaining useful life. Another method involves the use of deep learning to determine different equipment states. Ni *et al.* [349] used a Gated Recurrent Unit (GRU) RNN to predict bearing life over time. Bearing states are determined based on the bearing life until failure using the L10 calculation or predetermined vibration limits. Data-driven and physics-based models can predict and project the expected rise in vibration as a system degrades. Thereby, a manufacturer or maintenance engineer can determine when to schedule maintenance based on the most likely scenario.

2.6 Data Augmentation and Transfer Learning

Imbalance data in computer science refers to when certain classes from a dataset represent most of the samples over other classes. Depending on the distribution of the samples, the subsequent weighting of these different classes skews and affects the results of any data analysis applied. The common issue of data imbalance in condition monitoring stems from a lack of labeled data related to failure modes [22]. Often, in manufacturing, there is a prevalent amount of healthy data with varying label characteristics. Failures are sporadic in production, which is suitable for production. However, it needs more data for training condition monitoring applications. Kaur *et al.* [350] found 11 different application areas, which include Computer Vision, Medical Science, Image processing, and

Bioinformatics. Different approaches were taken in each domain to mitigate their effect on algorithm accuracy and robustness. This research work focuses on the impact of data augmentation and transfer learning to solve data imbalance. Data augmentation is artificially increasing the training data by modifying any existing data [351]. Transfer learning improves learning on a new task based on a related task [352]. These two methods are related to the data and knowledge for increasing learning for particular applications. DA and TL are two methods that fit within the domains for addressing condition monitoring approaches, such as preprocessing [353], hybrid methods, including ensemble learning and resampling [354], and algorithm-centered approaches [355]. The primary focus is due to the parallels this work draws with other domains in applying these two particular methods.

2.6.1 Data Augmentation

In computer vision and medical imaging, often training datasets can be skewed based on the number of available images within certain classes. For example, Spiesman *et al.* [356] recognized an imbalance of data related to certain Bumblebee species. To reduce the bias and variance from their model, they had to reduce the number of possible samples to represent each class equally. Another example is facial recognition in recognizing the bias from detecting people on watch lists [357]. It was found that commercial systems had issues guessing and determining a person's age as their skin became darker. Consider autonomous driving scenarios in which a system needs to tell the difference between a stop sign at night versus during the day [358]. In each scenario, the data imbalance from the data gathering can negatively impact the possibility of implementation in actual practice.

Data augmentation is one method to increase the amount of data for underrepresented data classes. In computer vision for deep learning, Shorten *et al.* [359] discussed different methods for changing images to represent different classes. Some of the proposed methods involve changing the orientation of the photo through rotation, flipping, translation, and cropping. Other methods consider changing the image filter by adding noise and changing the aspect ratio. Depending on the number of changes to a dataset, it can exponentially grow based on the number of changes to an image. There are few available images for training machine and deep learning algorithms in medical image datasets to determine diseases from medical X-rays and CT scans. Chlap *et al.* [360] considers many of the image augmentation techniques recognized by Shorten *et al.*. They also consider using Generative Adversarial Networks (GAN) to generate new images. GAN trains based on a set of images and then seeks to recreate these images based on noise input. The generated

images are then evaluated using a discriminator to determine their authenticity. In product quality for manufacturing semiconductor wafers, Saqlain *et al.* considered data augmentation methods to increase the training data for detecting wafer defects. In doing so, new images can be generated that mimic the configuration of old images. Data augmentation also extends to other fields. Liu *et al.* [361] considered a method of text data augmentation where certain words were either deleted, exchanged, inserted, or moved to increase the robustness of text processors. The number of sentences with the same meaning but changed words increases the robustness of a model meant to detect certain meanings.
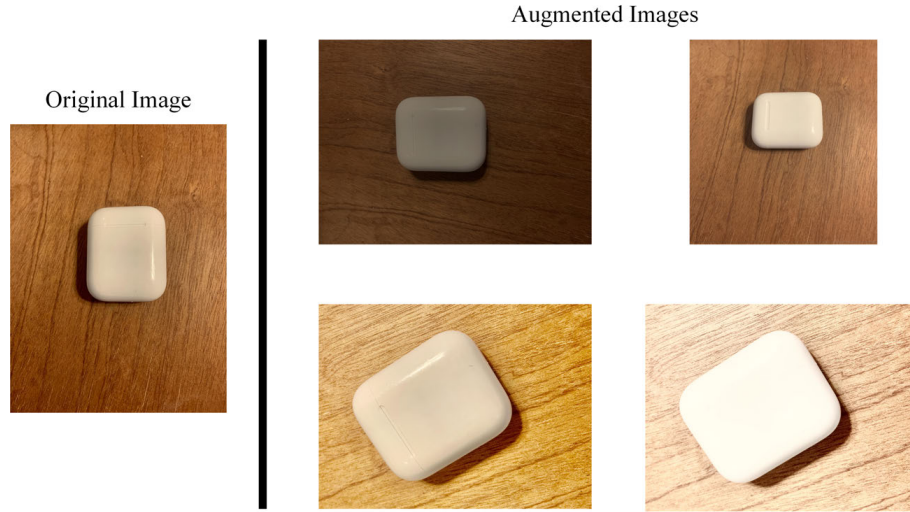
Figure 2.16: Example of different methods of data augmentation in detecting a case on a desk

In condition monitoring, there are usually two approaches for data augmentation: direct manipulation of the time series data or algorithmic-based augmentation. Direct time series manipulation can comprise similar methods as described in image data augmentation. One method is overlapping data samples to create additional sub-sample sets. Zhang *et al.* [362] suggest that creating over 50,000+ sample sets from 60,000 original data points was possible based on an overlap of 1 and a length greater than 2000. Li *et al.* [363] and Yu *et al.* [364] considered the use of noise addition, zooming in on the data, slicing the data and applying it to different segments, reversing the data, and amplitude shifting the data based on a scaling factor. Figure 2.17 shows an example of creating new vibration data based on local segment splicing. The rearranged and modified data can double the expected features based on the augmentation method.

In algorithmic-based augmentation, data augmentation uses either a Generative Adversarial
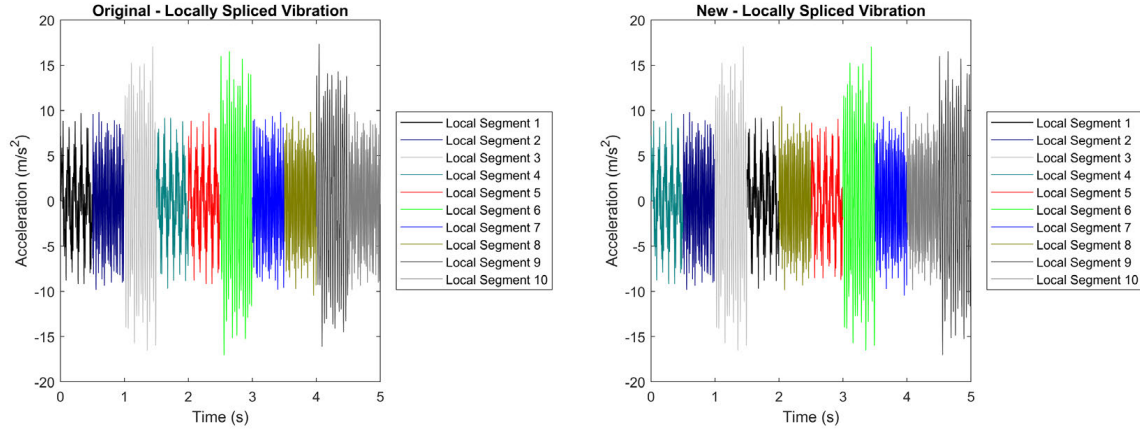
Figure 2.17: Example of time splicing vibration condition monitoring by splitting the original signal and rearranging it into a new signal based on local components

Network (GAN) or a Convolutional Neural Network (CNN) to modify and create new data. In addition to the data augmentation, these models diagnose faults using the true and augmented bearing data. Bui *et al.* [365] compared GAN-based data augmentation and that conducted by hand for FFT feature inputs. The generated GAN data could accurately predict the bearing fault above 90%; however, they performed 6% poorer when none of the original failure data were present. Shao *et al.* [366] considered using data augmentation with GAN to generate additional failure data. After the correction, none of the algorithms performed less than 99% when considering the balanced dataset. Wang *et al.* [367] utilized a GAN with an Autoencoder to diagnose bearing faults based on artificially generated data. The use of these models provides a streamlined platform for the generation of synthetic data.

2.6.2 Transfer Learning

In bearing condition monitoring, transfer learning considers data from the source data to contain the representative ratio of failure to healthy data expected in the industry. The target domain considers a scenario of insufficient bearing data to represent both classes. Both Schwendemann *et al.* [368] and Zhu *et al.* [369] combined CNN with transfer learning to inform better results in datasets where there were only a small amount of samples. They used the mean-max discrepancy to measure the source and target domain distribution. The measure ensures the distance on the space of probability remains similar. Other methods to measure the discrepancy between data distributions

are the Kullback-Leibler divergence and the Euclidean distance. Deviations from these values could signal non-convergence between the source and target domain. To demonstrate the adaptability of transfer learning, even in non-stationary conditions, Hasan *et al.* [201] created a CNN over variable working conditions to diagnose bearing faults. They use fine-tuning-based transfer learning, which transfers learned parameters to target models to save time during training and analyze bearing faults for different operating conditions. Zhao *et al.* [370] considers a feature-based learning method using manifold feature learning with dynamic distribution alignment to mitigate the potential for source and target drift in the data distributions. The technique used by Zhao *et al.* is unique as manifold feature learning is not a deep learning method, as was conducted by the previous applications.

As seen in the above-mentioned articles, several methods exist to conduct transfer learning and combinations to measure their success. Chen *et al.* [371] groups transfer learning methods into model-based, discrepancy metric-based, and domain adversarial-based methods. Model-based methods consider separate models that share learning parameters for transferring knowledge. Discrepancy-based approaches consider the same source and target domain model based on minimizing a discrepancy metric. Adversarial-based transfer learning is considered unsupervised and tries to modify the feature transformation rather than the model itself. Despite the successes with transfer learning, Chen *et al.* [371] present challenges concerning the interpretability of a deep learning model, the transferability assessment for different domains, and the reliability of data-driven models. Chen *et al.* [372] and Li [373] further echo these sentiments as challenges for transfer learning. Establishing additional criteria related to transfer learning would reduce the possibility of negative transfer between source and target applications. Figure 2.18 demonstrates the transfer learning process between two models for similar tasks. Model 1 represents a model trained using a large amount of labeled data. The parameters from model 1 are used as the starting parameters for model 2, meaning the knowledge transfer. Model 2 passes data with a low amount of labeled sampled data for training. However, using the parameters from model 1, the training process only needs a smaller amount of samples to complete the training process and reach a similar level of accuracy for model 1.
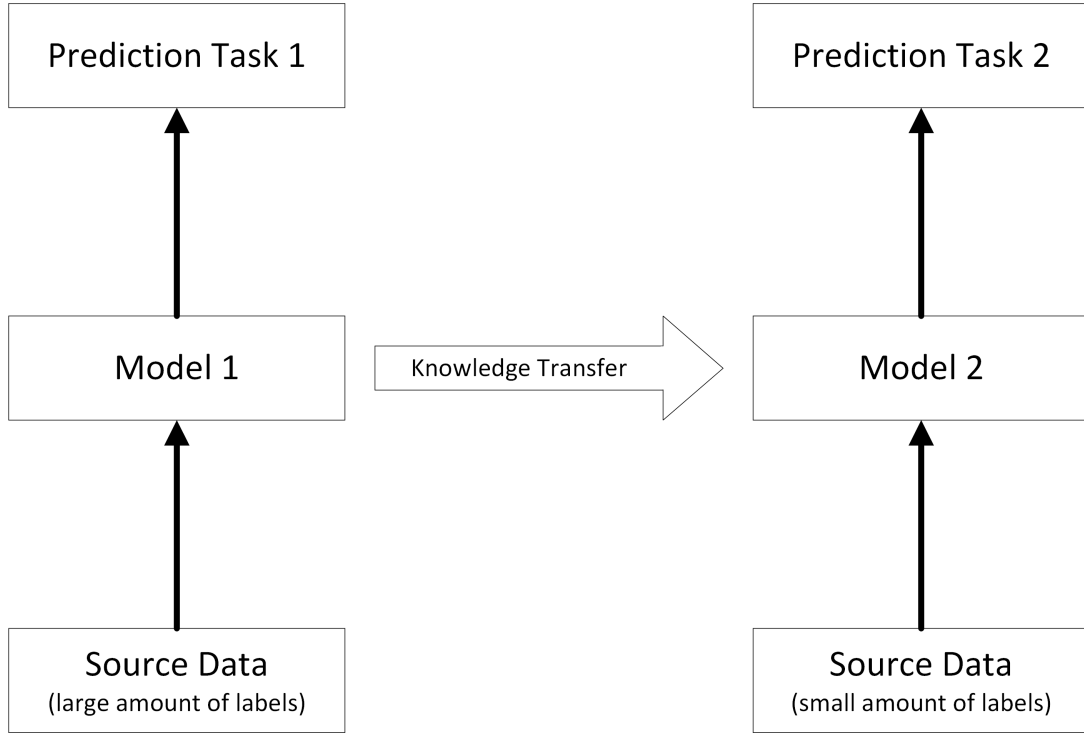
Figure 2.18: Example of transfer learning from Model 1 to Model 2

2.7 Digital Twins and Cyber-Physical Systems

Digital Twins (DTs) were first conceptualized by Grieves *et al.* [374] for product lifecycle management. Essentially, as one had their physical system, they would also have a digital system that mirrored that physical system. As changes were made to the physical design, changes would occur with the digital system and vice versa. Generally, though, the digital twin definition varies based on the user or organization. Shao *et al.* [375] did a comparison survey of different users for the digital twin characteristics by analyzing their definition, viewpoint, fidelity, and temporal integration. They found a wide variance across the four different definitions of a digital twin for condition monitoring, such as whether the temporal integration was in real-time or not and whether model fidelity was considered complete or partial.

For CM applications and PdM, the following definition is adopted: *A digital twin is considered a near real-time system that provides information related to the equipment's health to determine the appropriate time for maintenance schedule based on a combination of input data.* The physical domain encompasses all interfaces that work with the equipment, including the repair action

and data gathering. In contrast, the digital space considers all the data processing and analysis to generate information based on the inputs from the physical world. The phrasing allows for the type of model to be open-ended, provided it is either a data-driven, physics-based, or hybrid model [376]. Moghadam *et al.* [377] provides an example of a wind turbine physics-based digital twin using MSD equations to describe degrading components. Qin *et al.* [378] considers a data-driven model based on an improved CycleGAN to determine bearing fault diagnosis. The use of the model is to map data to the physical environment, and map features to the virtual environment to determine different bearing faults.

Cyber-physical systems (CPS) are systems that serve to integrate computational and physical processes [379]. CPS focuses on bridging the gap between the cyber and physical worlds. Using these systems, integrating new data and knowledge from either world to influence the other positively is supposed to be more accessible. Lee *et al.* [14] proposed a 5C architecture to describe the use of CPS at the component, machine, and production system level. The 5C architecture is considered to have the following attributes: Connection, Conversion, Cyber, Cognition, and Configuration, which increase in complexity from the connection to the configuration level. While DTs and CPS may appear the same on the surface level, there are inherent differences in their cyber-physical mapping and hierarchy, as a result, [380]. However, both are considered and introduced in this research, as they are used extensively in industrial applications. Furthermore, there is potential to use these techniques to package the acquisition, processing, and analysis techniques for bearing condition monitoring.

2.8 Background Synthesis

There is a key message to consider from each background section:

i.) **Bearings** are a common component that is used in a broad and diverse range of rotational equipment. Due to the application diversity, there is a wide variety of operating and environmental conditions one could consider in designing and selecting a bearing. These conditions could further influence the bearing performance and subsequent longevity in specific environments.

ii.) The **Cost of Unexpected Downtime** has far-reaching impacts outside of the equipment replacement. An unexpected failure could result in millions of dollars worth of damage for the

company from other cost sources, such as environmental, inventory, and labor costs. These costs are incurred in addition to repairing and replacing the equipment.

iii.) Maintenance personnel have created and integrated different *Maintenance Strategies* to keep up with evolving production lines and equipment. Due to the increased mechanization of products and processes, anticipating equipment failures for maintenance scheduling leads to lower costs in long-term planning but higher short-term costs for installation.

iv.) One of the common technologies integrated into manufacturing equipment is **Data Acquisition** systems to monitor different process values and equipment factors that may indicate progressing equipment failure.

v.) **Data Preprocessing, Processing, and Analysis** considers different techniques for cleaning and preparing data for analysis and the methods for transforming data into usable information. These methods are based on the statistical data change from the manufacturing equipment and the physics changes in the digital space.

vi.) **Data Imbalance** is a challenge when analyzing and predicting impending equipment failure when certain classes outweigh other classes, leading to skewed analysis results and inaccurate predictions. In the case of manufacturing equipment, the healthy class usually outweighs the failure classes, and in certain instances, no failure data is available.

vii.) **Digital Twins and Cyber-Physical Systems** are tools for encapsulating the data acquisition and analysis methods and interfacing with maintenance engineers. However, there is a loss of trust between manufacturers and these systems when inaccurate results are reported.

Chapter 3

Challenges and Solutions for Condition Monitoring Implementation

Chapter 2 provided a basic level of knowledge needed to understand the challenges for the presented solution in this work. In Chapter 3, the different challenges are framed about the different research objectives from the Introduction. The challenges listed in Research Objective 1 are to address the generation and labeling of data for condition monitoring systems. Research Objective 2 addresses the scaling and manipulation of data between different systems. Most of the research work is conducted in Research Objectives 1 and 2, validated in Research Objective 3.

3.1   Research Objective 1 Challenges

In manufacturing, as mentioned in Section 2.6, imbalance commonly stems from a lack of labeled samples concerning the different failure conditions of the equipment [350] in contrast to the healthy equipment modes. Hence, inadequately labeled data and low data availability can cause data imbalance. Supporting evidence mentioned earlier from Leukel *et al.* [23] that in some cases, the analysis is conducted with a dataset having less than 10 failures present in the sample population. In one such case, 30 million records (samples) were present, but only 8 failures. In other instances,

failure data is not recorded. Due to the lack of records, it is unclear how to create or apply any generated knowledge from prior lab work to actual production applications.

Fredriksson *et al.* [381], in observing and interviewing data scientists, listed common problems that contributed to data labeling stemming from the lack of a systematic approach, unclear responsibility for labeling, and noisy labels in the general data science domain. In the image labeling domain, to mitigate the effects of potential bias and variance in the data, researchers have introduced data labeling algorithms to reduce potential bias through crowd-sourcing [382] or bias algorithms [383]. However, manual data labeling through crowd-sourcing is time- and cost-intensive. Another challenge with crowd-sourcing data could lead to issues with label quality, depending on the knowledge level [384]. A common challenge in automated data labeling is to model and capture bias before implementing methods to remove it effectively [385]. If the bias algorithm is incorrectly configured, it may skew results to a certain conclusion and increase the prediction bias.

Consider the background sections from Section 2.1 through Section 2.5 as applicable background knowledge to frame the data imbalance problem for the data generation conducted in Research Area 1. There are various condition monitoring applications for motors [317], gearboxes [386], pumps [387], and turbines [388], and all are common applications in a manufacturing environment[1]. A common component in each listed application is bearings, which help control the rotational characteristics for these applications and others (*i.e.*, robots and CNC machine tool spindles). In each piece of equipment, bearings are listed as a common failure point in Failure Modes and Effects Analysis (FMEA), especially within industrial motors as [38, 39, 40, 37]. Due to their widespread application, condition monitoring applications focus on bearings as a primary point of failure using the methods described in Section 2.4 for the analysis methods from Section 2.5. Preliminary bearing data can establish a baseline, where any deviations could be marked for investigation or further analysis. As data is collected, it may or may not be assigned a label for further training data in a model. If failure training data is already available from early faults, a model may be deployed to monitor and identify potential equipment failures and predict future manufacturing life. As faults and errors occur along the production line, the program is adjusted for potential changes. This process outlines a brief overview of the condition monitoring implementation pipeline from ISO 17539 [5][2].

---

[1]These equipment are found in other industrial environments; the frame of reference for this work is in a manufacturing environment

[2]Further discussed in Section 4.1.2. Please reference Figure 4.3 for a detailed flow diagram for the general condition monitoring system

The CM system's configuration, upkeep, and data availability are where data imbalance affects this process. The alert, alarm, and failure criteria depend on existing data, information, and knowledge related to failures in the monitored system. If no failure data is available from the application, then the configuration relies on industry-expected limits, such as those from ISO [389]. These expected alarms and alert criteria may focus primarily on common failure modes and not consider process or equipment specific failure modes. Even when failure data is available for training, the data labeling procedure is hampered by the primary focus on returning the equipment to operate on the production line. Hence, significant data labeling may not occur, and there may be critical missing critical domain knowledge.

In addition, the labeling procedure changes based on the person implementing, collecting, and analyzing any CM analyses. For example, eight different bearing fault datasets are investigated in Section 4.1.1 to determine different characteristics for CM analysis. In each dataset, the data labeling, structure, and investigated failure modes vary to the point where eight different procedures yield potentially similar results. Hence, across CM literature, no standardized procedure or criteria is available associated with the data labeling and collection procedure to build these analysis tools to diagnose and predict bearing failure. The following references have noted the difficulty in collecting labeled data for condition monitoring applications:

  i.) [370] Zhao *et al.* note that it is difficult to gather the necessary labeled data for machine learning based on different operating conditions and equipment configurations.

 ii.) [390] Lyu *et al.* highlights some data-scarce regions when assessing product quality to inform operating performance.

iii.) [391] Cohen *et al.* state that industrial datasets are not always suited for supervised learning based on the need for extensive domain knowledge for accurate labeling.

 iv.) [392] Serin *et al.* finds that machine tool condition monitoring datasets require consistent labeling to ensure the proper metrics; otherwise, it invites unknown errors in the algorithm training process.

  v.) [393] Liu *et al.* note that it is somewhat considered infeasible to perform defect inspection while also labeling each sample collected from the manufacturing environment.

vi.) [363] Li *et al.* utilized data augmentation to create new data to circumvent the expensive cost of labeling data.

vii.) [394] Gao *et al.* noted that for big data implementations in manufacturing an increased amount of quality data is necessary to ensure prediction accuracy.

Data availability in condition monitoring systems is not necessarily an issue. Sensors and computational power are relatively low-cost for implementation along the edge or cloud computational. Data generation can occur at an exponential rate depending on the system and the number of systems under assessment. However, despite having access to the data, if it is not labeled or if there is no underlying information to determine what could have caused certain trends then the data does not provide a meaningful contribution to the understanding of equipment degradation.

3.2   Research Objective 2 Challenges

As data imbalance occurs due to low available labeled training data or failure data, there are methods to reduce the data imbalance by applying data techniques or algorithm-based methods. The background knowledge from Section 2.5 through Section 2.6 is applied to frame the problem regarding Research Area 2. Different data techniques to reduce data imbalance in condition monitoring consider reducing or creating training data amount to achieve a particular outcome. Sampling techniques, such as undersampling [395] and oversampling [396], can either add or remove data from certain classes to provide an even distribution of data for condition monitoring applications based on an existing set of data. However, duplicating data in oversampling can make any model overfitted [397]. By removing or reducing data for undersampling, data points are removed to reduce the imbalance between the different datasets [397]. However, removing that data could reduce the model performance in determining the correct diagnosis by affecting the variability within the system. Another technique to address data imbalance is data augmentation to create new data samples from an existing dataset.

As introduced in Section 2.6, data augmentation creates new data by applying modification factors, such as randomization, filtering, and scaling, to the original data set [398]. Depending on the dataset, specific data augmentation techniques will perform better than others depending on the data scenario[3]. Sometimes, it requires testing multiple data augmentation methods to improve results.

---

[3]Techniques are further explained in Chapter 5

The subsequent new data help to increase an algorithm's robustness to potential new data collected as the algorithm performs [399]. Data augmentation techniques have been applied in bearing condition monitoring to increase and improve datasets [363, 364]. Depending on the application, different techniques affect the algorithms tested. In addition, algorithms, such as Generative Adversarial Networks, create bearing data based on generators trained on failure data [400]. However, in this instance, these data augmentation techniques are based on an existing failure data source. Without failure data, the accuracy and applicability of data augmentation are reduced.

Pursuing an algorithmic approach, data imbalance can be solved by using tools, such as one-class classifiers [401], unsupervised learning approaches [329], and synthetic data generation [402]. In one-class classifiers, an algorithm trains on the status quo of bearing data and identifies potential outliers as defect points. Unsupervised learning approaches do not need labeled data and instead, attempt to classify data into clusters of similar data attributes to determine potential failures. Synthetic data generation with algorithms implies using a generator algorithm, such as a GAN. For example, a GAN algorithm is trained based on a set of failure data to generate additional failure samples that may resemble the original true data. These algorithms successfully predict equipment failures; however, the intrinsic knowledge needed to understand the identified outliers or clusters from the analysis is a drawback. In addition to data generation algorithms, an existing set of failure data is required to train the algorithm to generate the bearing data. It may be possible to extrapolate expected features from *similar* bearing data, but that may also skew results if they do not converge.

Another method to address data imbalance is using transfer learning to deploy similar but different algorithms to different scenarios. This process differs from the previously mentioned methods, as transfer learning considers data from two sources rather than the singular source in the prior techniques. For example, Wu *et a.* [403] designed a transfer learning method to train an algorithm based on generated laboratory data for practical, real-world scenarios. Wang *et al.* [404] used a mixture of ResNet-50 with a multi-scale feature extractor to transfer bearing data between applications to improve fault classification. However, an inherent issue with transfer learning is understanding how the algorithm tuning parameters change between different bearing applications. Generally, these different applications have different fault spectra that may not be present in the source or target domain, leading to negative transfer learning.

3.3 Digital Triplet Formulation

To summarize the previous sections (Sections 3.1 - 3.2), low amounts of or missing labeled failure data from inadequate labeling procedures cause data imbalance for manufacturing CM applications. Reducing data imbalance involves changing the data procedures, integrating additional steps in the data pipeline, or utilizing different analysis techniques. However, applying these other data imbalance mitigation methods does not guarantee increased accuracy or the ability to transfer generated data, information, or knowledge from one application to another. The lack of transfer power reduces the possibility of creating a generalized algorithm that fits the expected scenarios for bearing CM in a manufacturing environment. The time and effort related to labeling data from each bearing condition monitoring scenario lead to slower adoption for condition monitoring in a factory environment.

The Condition Monitoring Digital Triplet (CMDTr) is proposed to address the issues of low amounts of labeled data and how to transfer data, information, and knowledge between different applications in a manufacturing environment. Proposed initially as Surrogate Digital Triplet (SDTr) by Wescoat *et al.* [3], the CMDTr combines the work presented in this dissertation into three different areas to support digital systems in manufacturing. The three triplets proposed are the Physical Tiplet (PT), the Digital Triplet (DTr), and the Surrogate Triplet (ST). This system is meant to extend the DT and CPS discussed in Section 2.7. Generally, these systems are viewed as means to ingest data and return information [375]. However, knowledge of how data are transferred is sometimes lost in converting data to features and between the different applications.

Knowledge applications in digital twins have been somewhat previously investigated. Zhou *et al.* [405], who proposed a framework for a knowledge-driven digital twin manufacturing cell for planning, scheduling, and managing a manufacturing cell. In this scenario, the digital twin contains the information and knowledge to collect the manufacturing system. Knowledge-based systems have also been incorporated into digital shadows to characterize AI and human knowledge for manufacturing system operations [406]. Umeda *et al.* [407] previously introduced the digital triplet to characterize human interaction with a digital twin. In their interpretation, the digital triplet is meant to illustrate further the application of knowledge from the information generated in the digital twin aspect. With Zhou *et al.* and Ladj *et al.* [406], the knowledge is still encompassed in the digital twin to leverage the information readily.

However, one of the knowledge aspects that needs to be more readily captured in the current digital twin is the aspect of a generalized model that can be deployed to multiple systems of different sizes and configurations. A digital twin is considered specific to a physical counterpart; hence, the information and knowledge generated are meant for that particular application alone. The CMDTr intends to provide a method to extend knowledge to multiple systems using the Surrogate Triplet. The primary purpose of the CMDTr is to address the data imbalance problem in a digital twin scenario. A general understanding of the proposed condition monitoring triplet proposed is shown in Figure 3.1, showing how the data, information, and knowledge relationships reduce data imbalance and ease the implementation of condition monitoring systems in the manufacturing domain. The dotted lines represent the lines of knowledge from one triplet to another triplet, the dashed lines represent the change in data from one triplet to another, and the solid lines represent the change in information. In this current iteration, the condition monitoring triplet considers the knowledge from decision-making as another portion of the physical triplet. It is understood that some consider that the knowledge related to decision-making is part of the digital triplet in this scenario.
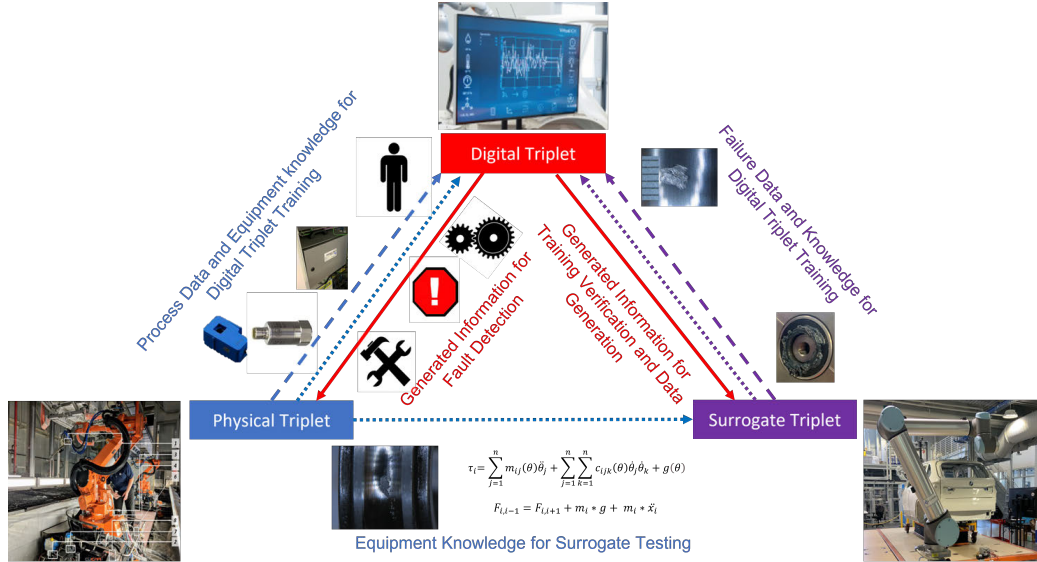
Figure 3.1: A general implementation of the Condition Monitoring Digital Triplet from [3]

This dissertation lays the framework for the necessary tools and methods to implement the CMDTr for bearings as depicted in Figure 3.2. As these tools and techniques are defined for bearings, the discussion is provided on how these methods are also extensible to other systems. In this research, Chapter 4 covers the initial data generation for the CMDTr to reduce gaps and holes

in the training profile for condition monitoring systems. This chapter demonstrates the methods and tools for designing the knowledge relationship between the physical and surrogate triplet and the data and information relationships between the surrogate and digital triplet. Chapter 5 considers the methods and tools to transfer the appropriate data between different bearing sizes and systems. This chapter will demonstrate the tools and methods for the knowledge relationship between the surrogate and digital triplet and the data and knowledge relationships between the physical and the digital triplet. Finally, Chapter 6 will provide the methods and tools to evaluate the effectiveness of new bearing systems using the generated data in training these condition monitoring systems. This chapter will explore the tools and methods for the information relationships between the digital triplet and the surrogate and physical triplets.
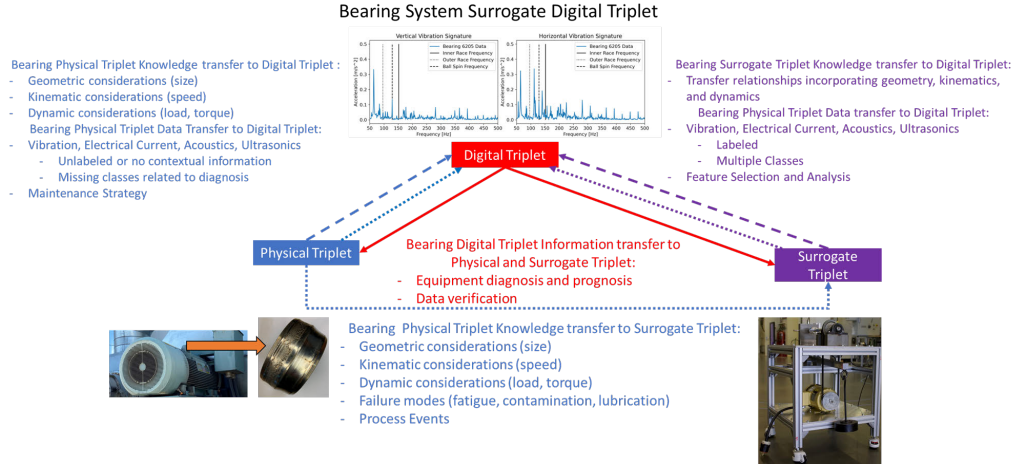
Figure 3.2: A general implementation of the Condition Monitoring Digital Triplet from [3]

Chapter 4

The Digital Triplet: Data
Generation

In implementing the CMDTr, the first consideration is determining the data sources for collecting data and ensuring the appropriate data amounts for diagnosing and predicting bearing failures. This chapter presents a purposeful failure methodology based on investigating bearing datasets but with applicability to other manufacturing equipment. Failure methodologies are crafted for bearings based on the associated failure modes defined by industrial standards, the equipment type, and the location within the equipment. The changes in the bearing failure signals are identified utilizing time and frequency analysis and compared to each other. As an additional consideration, non-stationary operating conditions are considered in the bearing data.

4.1 Purposeful Failure Methodology

The *Purposeful Failure Methodology* (PFailM) was created as a result of the investigation into RQ1.1 and RQ1.2:

i.) Research Question 1.1: How are the existing bearing datasets deficient for training manufacturing condition monitoring systems?

ii.) Research Question 1.2: What methodology criteria are necessary to fill gaps for condition monitoring datasets utilizing a surrogate triplet?

During the investigation, standardized criteria of data attributes are applied to eight publicly available bearing datasets. From these criteria and their subsequent ranking, gaps are identified with condition monitoring datasets as possibilities for improvement for standardizing data generation in condition monitoring. A review of design, failure, and condition monitoring implementation methodologies combined to form the steps necessary for the PFailM.

4.1.1 Condition Monitoring Data Assessment

Open-source datasets help improve machine learning algorithms by providing more areas of learning for the related systems. Examples of open-source datasets include the MNIST handwritten images and the ImageNet dataset in the computer vision domain. These datasets provide a variety of different representations for certain labels. Using these datasets provides data scientists with hundreds of thousands of training samples that could potentially increase with techniques such as data augmentation and oversampling. The open-source nature allows for ease of access in utilizing these datasets in varying scenarios. Popular sites that host open-source datasets are places like GitHub[1] and Kaggle[2]. Other organizations, such as IEEE[3], Mendeley[4], and NIST[5], will also host open source datasets for various domains and projects.

For condition monitoring datasets specifically, a detailed list of complete condition monitoring datasets can be found on the NASA Prognostics Center of Excellence[6], IEEE Dataport, Kaggle, and GitHub. The CM datasets available consider multiple different systems, such as bearings, gearboxes, composites, and electrical equipment. Each dataset provides a different level of contextual knowledge in terms of documentation. Hence, there are varying levels of how to utilize the available data for condition monitoring. Hence, an initial survey of condition monitoring datasets was considered to determine what criteria are commonly found in these condition monitoring datasets. At the end of Section 4.1.1, a ranking and review of the current data dimensions is provided.

Since the primary focus of this dissertation is bearings, the initial data survey considered bearing datasets from eight different sources. These bearing datasets are hosted on university websites, third party organizations, such as NASA and IEEE, or open-source data forums such as

---

[1]Link to the front page of GitHub: https://github.com/
[2]Link to the front page of Kaggle: https://kaggle.com/
[3]Link to IEEE Dataport: https://ieee-dataport.org/
[4]Link to Mendeley Data: https://data.mendeley.com/
[5]Link to NIST datasets: https://www.nist.gov/el/ammt-temps/datasets
[6]Link to NASA Prognostics Center of Excellence: https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository

GitHub and Kaggle. Table 4.1 lists the publicly available data for bearings used for this assessment by providing a data overview for the collection location, associated publication, data types, test conditions, and dataset size. In some research works, the IMS bearing dataset refers to the NASA bearing dataset since it is hosted on the NASA Prognostics Center of Excellence (PCoE) (note however that the NASA PCoE also hosts the IEEE 2012 PHM challenge dataset). Each dataset is, to mitigate potential confusion, referred to by the original collection entity to remove any potential naming confusion. Each dataset has been publicly available in the last 10 years except for the IMS dataset, which has been available for over 20 years. The datasets are hosted by the NASA PCoE, GitHub, Mendeley Data Center, or their respective university web pages and are verified by a corresponding journal citation[7]. The bearing datasets are referred to based on the agency that performed the data collection. The bearing test methods are described based on the methods that faults were induced in the bearings. The acronyms mean the following: O-AF (Overload - Accelerated Failure), S-AF (Seeded - Accelerated Failure), and SF (Seeded Faults). The difference between these failure methodologies is further described in Section 4.1.3. The data type acronyms mean the following: Vib (Vibration), Temp (Temperature), and EC (Electrical Current). These datasets are not the only available bearing datasets, but they have been used extensively in bearing condition monitoring research. In addition, other types of condition monitoring datasets exist; however, bearing datasets provide the most breadth in terms of availability and quantity.

Table 4.1: Publicly available bearing data

Bearing Datasets	IMS	FEMTO	CWRU	MFPT	UO	PU	XJTU-SY	KAU
Associated Reference	[408]	[409]	[410]	[411]	[412]	[413]	[414]	[415]
Bearing Experimental System	4 Bearing Test	1 Bearing Test	Motor Bearing Test	N/D	1 Bearing Test	1 Bearing Test	1 Bearing Test	1 Bearing Test
Bearing Test Methods	O-AF	O-AF	SF	SF	SF	O-AF, S-AF	S-AF	SF
Failure Modes Data Types	Fatigue Vib (2-axis) / Vib (1-axis)	Fatigue Vib (2-axis), Temp	Fatigue Vib (1-axis)	Fatigue Vib (1-axis)	Fatigue Vib (1-axis)	Fatigue Vib (1-axis), EC, Temp	Fatigue Vib (1-axis)	Fatigue Vib (1-axis)
Data Amount	6.08 gb	2.97 gb	0.709 gb	0.057 gb	0.762 gb	21.28 gb	4.23 gb	0.001 Mb

---

[7]Links to the Bearing datasets are found in Table C.1 in Appendix C

The quality of the dataset is assessed and checked using a standardized criteria. Lessmeier *et al.* [413] proposed a criteria for reporting bearing data in their dataset formulation based on ISO 15243 [6]. In the criteria, they report the test length, damage methods, and operating conditions, which are a part of bearing condition monitoring investigations. In a separate analysis, Hagmeyer *et al.* [416] considered criteria for dataset assessment to represent data scenarios for data generation. In these data scenarios, they discuss the methods of data history range, acquisition, the damage degradation, system information, data quality and quantity, and data source. In each avenue, these scenarios provide the measures for describing the datasets as needed for PHM datasets; however, not all of the criteria are transferable to the analysis or extensible to other similar bearing systems. However, they are good measures of how to describe a system.

Considering the above mentioned criteria a more simplified approach was taken for describing dataset configurations. Hagmeyer *et al.* and Lessmeier *et al.* provided a background for how to view condition monitoring datasets and ways to describe the dataset However, a more general criteria was sought that could provide a more critical analysis of prior condition monitoring data assessments with general data criteria. In 2002, Pipino *et al.* published a set of data dimensions describe data from businesses looking to leverage more analytics in their decision making. The criteria, also referred to as data dimensions, were originally derived based on work conducted by Wang *et al.* [417] in a data quality review from the prior research literature that synthesized the above attributes. The data dimensions also tie into work deriving information quality dimensions conducted by Kahn *et al.* [418]. The focus remains on *data* quality dimensions rather than *information* quality dimensions. The data dimensions from Pipino *et al.* describe data attributes that ensure the data accurately represent the conditions and classes by encompassing the necessary labels to trace and transfer data between similar applications.

For the analysis, the dimensions considered by Pipino *et al.* were split into an analysis of either the data or dataset. The split was conducted because it was considered that some of the dimensions described the overall dataset rather than the actual data. Table 4.2 contain the data criteria described by Pipino *et al.* that was determined to pertain directly to the data. Table 4.3 contain the data criteria described by Pipino *et al.* that was determined to pertain directly to the dataset. Some of the dimensions were also combined as they provide overlapping definitions. Based on the split of data and dataset dimensions, there are overlapping instances with the work conducted by Hagmeyer *et al.* and Lessmeier *et al.*. However, the underlying dataset considerations was not

considered in their work and is considered novel versus prior work in this area. In the following sections (Section 4.1.1.1 - 4.1.1.3), the dimensions are described how they asses the respective bearing datasets and determine whether these datasets are good for CM applications.

Table 4.2: Data Quality Dimension from Pipino *et al.* [2] and the interpretation for Condition Based Monitoring Datasets. These dimensions describe the data.

Dimensions	Definitions by Pipino *et al.*[8]	Definitions interpreted for Condition Based Monitoring Datasets
Appropriate Data Mount	Extent to which the data volume [are] appropriate	Determine the dataset sample count across the different operating conditions and failure modes in relation to equipment failure diagnosis
Believability/ Free-of-Error	Extent to which data [are] regarded as true and credible/ Extent to which data [are] correct and reliable	Assess the accuracy of data labeling and supporting information for the different data classes before analysis/ Determine if the fault and operating characteristics are present in the data based on inspection
Completeness	Extent to which data [are] not missing and [are] sufficient breadth and depth	Determine if there are any inaccuracies in the data representation
Consistent Representation/ Ease of Manipulation	Extent to which the data [are] presented in the same format / Extent to which data [are] easy to manipulate and apply to different tasks	Determine if the data are organized the same way for each data collection and test / Assess the different possible methods for translating data to features based on the data organization

4.1.1.1   Data Dimensions

Appropriate Data Amount

The appropriate data amount is not accurately defined for training condition monitoring applications. Leukel *et al.* [23] surveyed that industrial applications can have less than 10 failures in their training data. Hence, the interpretation of the failure training data size varies depending on the application and analysis method. The determination of dataset size is a common problem for data analysis. In some cases, a study's uncertainty and confidence intervals may determine the amount of data needed [419]. Hackshaw [420] theorized the data amount needed is based on the study objective. The larger the sample size in the study, the more precise the conclusions for the

Table 4.3: Dataset Quality Dimension from Pipino *et al.* [2] and the interpretation for Condition Based Monitoring Datasets. These dimensions describe the dataset.

Dimensions	Definitions by Pipino *et al.*[9]	Definitions interpreted for Condition Based Monitoring Datasets
Accessibility/ Timeliness	Extent to which data [are] available/ Extent to which the data [are] sufficiently up-to-date for the task	Assess whether the datasets are publicly accessible for research and public use/ Assess how up-to-date the data are for their respective data collection
Concise Representation	Extent to which data [are] compactly represented	Assess how well the data are condensed for analysis
Interpretability/ Understandability	Extent to which data [are expressed] in appropriate languages, symbols, and units/ Extent to which data [are] easily comprehended	Ensure that the units are properly labeled for the data/ Assess the data descriptions for the dataset to improve data understanding
Objectivity	Extent to which data [are] unbiased, unprejudiced, and impartial	Determine how repeatable the condition monitoring data generation methods
Relevancy/ Value-Added /Reputation	Extent to which data [are] applicable and helpful/ Extent to which data [are] beneficial and provides advantages from its use/ Extent to which data [are] highly regarded	Assess how the data are labeled concerning expected system failure modes/ Assess how well the dataset generates and adds new data to a CM system / Assess dataset impact for condition monitoring and evaluation

study. Small sample studies achieve short-term goals, but may not apply to larger populations unless through added scalability or uncertainty.

Due to their widespread application, bearings fail at varying rates and for varying reasons. However, as noted by major bearing companies such as SKF [7], bearing failures are linked to four failure causes: lubrication, contamination, assembly/operating errors, and general fatigue. The failure causes induce different failure types depending on the bearing application, where it occurs, and how the damage progresses as listed in ISO 15243 [6]. Bearing research has investigated these failure causes and effects by adding contamination into the lubrication [87], starving bearings of lubrication [421], placing artificial dents to induce bearing mounting defects [422], and the overloading the bearing on inducing early-stage fatigue damage [408]. The weighting of the appropriate amount of failure data varies based on the application and the likelihood of one failure occurring over another.

The appropriate data amount is considered based on the number of potential failure classes

from the datasets due to the general lack of standardized metrics for evaluating the data amount. Table 4.4 shows the breakdown of scoring for the appropriate amount of data for the bearing datasets. Four dataset characteristics were determined to compare the data amount between these datasets: Failure Modes, Operating Conditions, Damage States, and Applications. The failure causes evaluations to consider the different methods for inducing damage in bearings to gather data. The operating conditions and damage states consider the extent of failure tested for each dataset. The application row considers the different failure test beds used. The raw score is shown in the total row. In the dataset ranking, these scores are normalized and scaled to a weighting of 10, where the highest score receives a 10 and rounded down.

Table 4.4: Total data amount ranking for the bearing datasets

Bearing Datasets	IMS	FEMTO	CWRU	MFPT	UO	PU	XJTU-SY	KAU
Failure Causes (4)	1	1	1	1	1	2	1	1
Operating Conditions (3)	1	2	5	8	4	4	3	2
Damage States (2)	4	17*	15	3	4	3	15*	2
Applications (1)	1	1	1	4	1	1	1	2
Total	16	45	50	38	25	27	44	16
Normalized Score	3	9	10	7	5	5	8	3

From this analysis, the CWRU dataset performs the best based on the consideration of the different damage states tested and the operating conditions. The primary reason for the breadth of data generated stems from having the second most operating conditions and damage states tested. The XJTU-SY and FEMTO dataset perform equal to or greater than the CWRU bearing dataset for the number of damage stages tested but fall short with consideration for the operating conditions. The asterisk for the XJTU-SY and FEMTO datasets' damage states comes from an assumption that each bearing tested provides a unique damage stage, as their documentation did not clearly define the different damage states in their run-to-failure testing. This data dimension favors datasets with a wider breadth of data generated versus only considering the data amount since the amount needed for condition monitoring applications is subjective. Future analysis could consider the overlap between the data amount needed as a function of the failure modes and potential failure breadth.

Completeness. The data completeness was conducted to verify that the data contained the right amount, as stated in their documentation. Incomplete data refers to instances where the reported

data amount does not match what the dataset contains. Missing data is a common problem in condition monitoring due to the equipment operating environments. Methods are available to handle missing data by approximating the data or removing those data from the sample set [423]. However, if these occur at regular intervals, it could speak to more problems with the data acquisition system and inhibit any real time analysis. Hence, it is an important data characteristic to note in any training dataset.

Only the IMS dataset had significant amounts of missing data for the datasets assessed. Figure 4.1 shows an example of Test 1 missing data for the IMS dataset through the representation of the full test in Figure 4.1 (a) to what is meant to happen as shown in Figure 4.1 (c). The missing gaps could be attributed to errors with the data acquisition process or errors in testing. It could signal that there is potentially missing information from the tests that may confuse further analysis. It should be noted that Figure 4.1 (c) shows a change in the file frequency that was called out in the IMS documentation. However, it does not explain the missing files in other parts of the dataset.
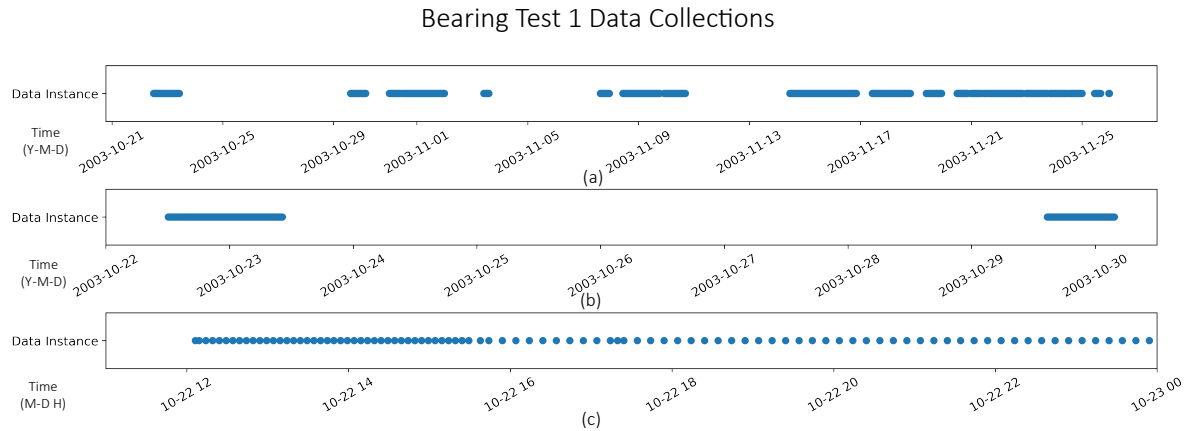
Figure 4.1: Missing Bearing Test Files and Sample Rate Verification from IMS Test 1. (a) shows the change over a period of days, whereas (b) shows the missing data for one section zoomed in over hours. (c) shows the change in data accumulation as and the instances of too many time recordings in the data.

The other datasets did not have larger missing samples. The other datasets had, at most 1%, of their data missing from their particular files. The percentage of data missing was calculated based on the number of samples, the sampling rate, and the sample duration. In the case of the FEMTO and CWRU datasets, there were instances where the files were present, but data were missing from columns in those file categories. The other datasets had negligible missing data based on the comparison of missing data found in the IMS datasets. Hence, the amount of missing data

from the IMS, FEMTO, and CWRU was assessed against the amount of expected data and used to compute their score relative to a weighting of 10. The dataset weighting was determined starting at 10 and then decreased based on the amount of data missing as a percentage from each dataset.

Believability/ Free-of-Error. The believability dimension determines if the necessary class attributes are present in the data based on their labeling. In contrast, the free-of-error dimension assesses whether the necessary signal attributes are present. For bearing condition monitoring, the believability dimension determines if the equipment operating information is present, the fault characteristics are properly labeled, and any additional supporting equipment characteristics are recorded. Equipment characteristics correspond to the bearing size and dimension information, which will influence the data generation. Table 4.5 details whether each dataset includes the corresponding information for the labeling and the supporting information for the different data classes. There are three critical aspects for fault labeling: damage location, damage size, and bearing lifetime. The datasets use terms such as "outer race", "rolling element", or "inner race" describe the location of the defect. Combination faults are where multiple bearing elements are damaged. For example, a combination fault could include damage to the outer and inner races for the same bearing. All three elements could also have damage as well. It is not noted in any of datasets whether any of the datasets have multiple rolling element defects. The damage size, when reported, could include the defect size's length, width, and depth. In some cases, the bearing life will not always be reported. Bearing life measures how much damage has accumulated over time with the bearing or the test-stopping point. When bearing life is recorded, it can be given in terms of defect size or time since the operation started. Each class has a speed label associated with the operating conditions to account for the bearing speed. The speed is important for calculating bearing life using the L_{10} equation by converting the number of revolutions into time and frequency information to determine the absolute bearing characteristic frequencies. None of the datasets specifically record the bearing load limits; however, it is noted that this information is readily accessible based on the bearing manufacturer. The bearing load and torque during operation are recorded based on the dataset. Recording both or either influences the resulting bearing vibration amplitude. The bearing size and dimensions were the common elements relative to equipment characteristics among all the datasets, except in the case of the FEMTO bearing dataset. Not included in the category scoring, but noted in this dimension, is whether there is photographic evidence to quantify faults. The KAU, XJTU-SY, PU, FEMTO, and IMS included pictures, but those were to provide examples of specific

79

faults. The score was based on a total of 10, with a weight of five for the fault labeling, three for the operating conditions, and two for the equipment characteristics. The scores were based on the number of possible yes's in each category. As long as the load or torque was recorded for the operating conditions, it was considered a full score with the associated speed.

Table 4.5: Criteria for describing bearing labels

Bearing dataset	Fault Labeling	Operating Conditions	Equipment Characteristics	Pictures Related to Faults
IMS (NASA)	Damage Location: Yes, Damage Size: No, Bearing Lifetime: Yes	Speed: Yes, Load: Yes, Torque: No	Yes	Yes, but only example faults
FEMTO (IEEE 2012 PHM)	Damage Location: No, Damage Size: No, Bearing Lifetime: Yes	Speed: Yes, Load: Yes, Torque: No	No	Yes, but only example faults
CWRU	Damage Location: Yes, Damage Size: Yes, Bearing Lifetime: Yes	Speed: Yes, Load: No, Torque: Yes	Yes	No
MFPT	Damage Location: Yes, Damage Size: No, Bearing Lifetime: No	Speed: Yes, Load: Yes, Torque: No	Yes	No
UO	Damage Location: Yes, Damage Size: No, Bearing Lifetime: No	Speed: Yes, Load: No, Torque: No	Yes	No
PU	Damage Location: Yes, Damage Size: Yes, Bearing Lifetime: Yes	Speed: Yes, Load: Yes, Torque: Yes	Yes	Yes, but only example faults
XJTU-SY	Damage Location: Yes, Damage Size: No, Bearing Lifetime: Yes	Speed: Yes, Load: Yes, Torque: No	Yes	Yes, but only example faults
KAU	Damage Location: Yes, Damage Size: Yes, Bearing Lifetime: No	Speed: Yes, Load: Yes, Torque: No	Yes	Yes, but only example faults

In contrast to the believability dimension, the free-of-error dimension for bearing condition monitoring focuses on the different analysis techniques to verify the bearing data. Clean data, in this sense, is described as data that have little to no noise that could hide or confuse any of the healthy signals [424]. The expectation for clean data is not a realistic one; however, there are specific indicators to differentiate the deviation between expected healthy and failed data. For example, rotational equipment, such as bearings, have particular characteristics in the frequency spectrum based on the fault modes, equipment configuration, operating parameters, and bearing size. Table 4.6 contains notes on verification methods used for each dataset or attached research documentation. This analysis considers only the methods used by associated publications. The IMS bearing and FEMTO datasets were the only two datasets that used time-based and frequency-based features

described earlier to verify the damage characteristics for at least a portion of the data. The CWRU, MFPT, UO, PU, and KAU data used frequency-based features to identify the bearing characteristic frequencies to identify bearing damage. The XJTU-SY bearing data includes no data features to verify or validate damage. However, they demonstrate the raw data change from a baseline to a fault condition. In some instances, different machine learning techniques were presented, such as in the case of the PU and XJTU-SY datasets, to demonstrate the separability among the data classes. Due to the use of both time techniques, the FEMTO and XJTU-SY datasets score a 10, whereas the other datasets only receive a 5. The weighting was determined based on the number of possible outcomes. Since there were only two, datasets received a score of a 5 or 10.

Table 4.6: Metrics used to verify bearing damage

Bearing Dataset	Time-Based Features	Frequency-based Features
IMS (NASA)	Verifies data from Test 1 with RMS and Kurtosis measures to determine if the damage is present	Utilizes wavelet analysis to extract early defect variation in the data for identification of particular bearing faults
FEMTO (IEEE 2012 PHM)	Verifies data from specific tests using the Crest factor and K-factor in their verification	Uses spectrogram analysis to view how the power spectral density changes concerning time
CWRU	Not Provided	Verifies the data using three different frequency transform methods utilizing Envelope Analysis using the standard method, with cepstrum prewhitening, and a discrete benchmark method to extract bearing characteristic frequencies
MFPT	Not Provided	Verified using Envelope Analysis for different fault conditions to extract bearing characteristic frequencies
UO	Not Provided	Verified using Time-Frequency extraction to verify the damage bearing characteristic frequencies
PU	Not Provided	Verified using Envelope Analysis to extract bearing characteristic frequencies
XJTU-SY	Not Provided	Not Provided
KAU	Not Provided	Verified using Fourier Analysis to extract damage characteristics

Consistent Representation/ Ease of Manipulation. The consistent representation dimension assesses how well the data are organized and if it is compatible for each file. CM systems require a standardized procedure to ensure minimal data uncertainty when performing equipment health analysis [71]. The same file organization and data collection procedures during experimentation minimize data uncertainty by ensuring that the data files are formatted the same for ease of access. Consistent formatting allows for efficient data and feature extraction and expedited processing. Three different file structures are considered for the bearing datasets: .csv files for the FEMTO and XJTU-SY data, a text-file variant for the IMS data, and .mat files for the UO, PU, CWRU, MFPT,

and KAU data. Each file type in Python has a unique command to access the data content. If the data files are each organized in the same format, a singular command can extract the data from each file without needing personalized file calls. An initial file was loaded for each dataset to understand how to access the data and formulate a function to check that each file is read. The functions were unique to each dataset and written using Python v3.8. Each function was able to access the data within each file. There was one individual file in the PU dataset where the PU function did not work due to a column misrepresentation. The data from the PU file was manually extracted to account for the data. Since it was only one file out of all the tests conducted, only a minor penalty (loss of a point) was assessed against the PU dataset. Consistent representation is typically an established norm as the original publishers of the data would want to ensure it remains in a format for easy analysis.

The ease of manipulation dimension refers to how well the dataset is structured to facilitate changing the data into features for analysis. In this dimension, the focus is primarily on feature extraction techniques for bearings. Feature extraction for condition monitoring refers to reducing the data dimensionality to a set of features analyzed to diagnose equipment failures [425]. Feature extraction is used interchangeably with feature generation; however, this work differentiates the two by describing feature generation as the method used to transform data into features [273]. For bearings, features are generated in either the time or frequency domain. Time domain features refer to values related to statistical changes in the signal, such as monitoring increases in the root-mean-square (RMS) or increasing variance in the gathered data [426]. Frequency features refer to values that occur at specific instances, such as the rotational frequency of the equipment and the bearing fault frequencies [296]. The frequency features are commonly used in the rotational analysis as a physical representation of the damage present in the equipment. For bearing condition monitoring, several frequencies denote particular condition criteria related to the overall equipment health, such as the operating speed, system misalignment, and mechanical looseness.

For the completeness dimension, only the PU dataset had issues accessing each data set similarly. The scoring was based on a percentage of correctly accessed files relative to the incorrectly accessed files. That percentage was then multiplied by ten and then rounded down. Each dataset subsequently scored 10, except for the PU dataset at 9. For ease of manipulation, each dataset allows for the conversion of raw data into features easily. However, the dataset collection strategies allow for additional analysis to take place. In the case of the IMS, PU, FEMTO, and XJTU-SY

datasets, they are collected concerning time during the test duration. The CWRU dataset contains data concerning different damage stages with a failure case. The UO, KAU, and MFPT datasets are collected based on each failure case and do not concern time or increasing damage. The progressive collection allows for testing about time and progressing damage. Hence, the IMS, PU, XJTU-SY, and CWRU datasets receive a ten as they allow for feature generation related to diagnosis or prognosis analysis. The MFPT, KAU, and UO datasets received a five as the data comparison only occurs concerning baseline and failed cases. The weighting was determined based on the number of possible outcomes. Since there were only two, datasets received a score of 5 or 10.

4.1.1.2 Dataset Dimensions

This section provides a brief overview of the other data dimensions provided by Pipino *et al.* [2]: accessibility, concise representation, interpretability, objectivity, relevance, and security. The accessibility dimension establishes whether the data are available for public and research use in training condition monitoring applications. Due to the data quantity in circulation, restrictions are often placed to protect the originators' rights regarding data use for outside entities using legal frameworks [427]. These restrictions are codified using Non-Disclosure Agreements (NDA) or company policies to prevent private data from reaching the public domain where it may harm the company or business practices. For this review, the accessibility dimension is framed toward assessing which bearing datasets are known and publicly available. Each dataset is stored either in a public site such a hosting sites, university web page, or government organization. Other condition monitoring data are available on similar sites associated with datasets describing gearboxes, motors, batteries, and other electrical components. The benefit of utilizing these datasets over data from proprietary organizations is the unrestricted data use with appropriate citations. The bearing datasets in the review are all equal for the accessibility dimension due to their public location. There are other locations where these datasets may be stored with parenthetical information corresponding to other names for the datasets. However, the risk of not providing equal access to data reduces the potential for reproducing research. The timeliness data dimension refers to how up-to-date the data are with respect to the documentation and the point-of-contact for questions. The timeliness dimension considers only when the data are updated.

The concise representation of the dataset refers to how well the data are packaged and condensed to reduce the file size for storage and processing. Concise representation is considered for

textual analysis due to the large amount of words that accompanies any text processing. Concise representation in text analysis focuses on how to prioritize sections of text documents to improve text analysis [428]. For this analysis, the datasets are considered concise based on the storage size each dataset contains. The general structure for condition monitoring data can contain a range of additional information that may or may not pertain to the bearing condition. The extraneous data may reduce the processing rate possible for the bearing data during fault analysis. The comparison in for this metric is to get the average collection size for each bearing dataset represented in a MATLAB file type. A MATLAB file type was chosen because the file type is the most common file format for each dataset. After the comparison of only raw data, a second comparison was considered by removing all of the extraneous data except for the vibration values. Only the vibration values were considered since these are the similar data types between the different sources.

The interpretability dimension assesses whether the data have the correct contextual labels for inferring the data representation. Vibration can be expressed in three different units: displacement (mm), velocity (mm/s), and acceleration (m/s^2). The understandability data dimension assesses how well the data are described through their data documentation. Across the dataset documentation, similar document sections describe the methodology for generating data, the file structure for accessing the data, and the dataset structure into the individual classes.

The objectivity dimension determines how repeatable the data are under different operating conditions and failure modes. Objectivity in data focuses on reducing the bias that may influence a decision. Bias in the data can cause errors when one class is weighed favorably over another class thereby potentially skewing results [429]. Having untreated bias in algorithms and data in condition monitoring can cause failure misdiagnosis and possible subsequent unplanned downtime. For the objectivity dimension, the number of tests conducted and the number of bearings that failed for each test are considered.

The relevance dimension assesses how applicable the data are in identifying the specific labeled failure modes. The reputation dimension focuses on the impact of the dataset in condition monitoring training and validation. In transportation systems, poor data quality or ingestion approach causes poor decision-making, leading to potentially life-threatening conditions. To that end, researchers, such as Chuprov *et al.* [430] and Acharya *et al.* [431], have looked to incorporate data quality metrics and data management systems to assess data before use. Including a reputation factor with each data set creates an associated confidence value about the potential information

84

generated. The number of downloads has been considered for dataset assessment as a reputation metric; however, that metric is not tracked or publicly available across all possible hosting sites. The value-added dimension assesses how well the training data adds new information to a CM system. The daily data generated from a condition monitoring system can overwhelm a processing system if not correctly managed. Hence, a criterion is needed to determine what additional data is needed outside of the initial data collection.

The data security dimension assesses what data restrictions are in place and protection methods are used to prevent malicious modification. Because these datasets are public, there are no associated restrictions for their use in the research literature as long as the authors provide the appropriate citations and references. There are instances where partial versions of each dataset are available from third-party users. In the data dissemination world, partial or even malicious data provides as much damage as no data. Hence, the adoption of data security frameworks should be considered for future data dissemination to ensure data quality [432, 433].

These dataset dimensions are not incorporated into the final scoring for the dataset ranking since many of these dimensions are corrected in the post-processing any collected data. The other criteria from Pipino *et al.* [2] are considered more dataset-level criteria. In this analysis, dataset-level criteria are associated with the methods in how the data are stored and disseminated, rather than in the acquisition and analysis phase. An example is the concise representation dimension, where the collected data are assessed based on how well the data are packaged for analysis and processing. The file type, the number of significant figures, and the data type during the analysis influence the data size. Hence, the dimension does not directly impact the data acquisition for condition monitoring, as this dimension can change based on the user storing and accessing the data for future use. A similar dimension consideration is the security dimension. In open-source medical datasets, data are anonymized before dissemination to protect patient identity [434]. Similarly, security does not necessarily impact the dataset formulation until the preparation for dissemination. Future work could focus on these dataset dimensions for the optimization of data processing and analysis, especially for real-time analytics.

4.1.1.3 Bearing Dataset Review and Ranking Relative to the Data Dimensions

Table 4.7 contains the final dataset ranking score relative to the different sub-scores from the analyzed dimensions. Based on the dimensions, the datasets are grouped into two distinct

categories. The IMS, MFPT, UO, and KAU datasets had scores within 3 points of each other across the different categories and centered at 40. Separately, the CWRU, FEMTO, PU, and XJTU-SY datasets were also within 4 of each other and centered at 51. The two highest scoring data sets, the CWRU and FEMTO datasets, maintained the highest or second highest score across at least three of the six categories. Across the different categories, the Appropriate Data Amount and Believability dimensions differentiated the data set the most across the different categories. A part from that the datasets that did not perform as high were usually deficient in at least two of the other four categories.

Table 4.7: Dataset ranking based on the scores

Bearing Data Set	A.D.A.[1]	Completeness	Believability	F.E.[2]	C.R.[3]	E.M.[4]	Total
IMS	3	7	8	10	10	10	39
FEMTO	9	9	4	10	10	10	52
CWRU	10	8	10	5	10	10	53
MFPT	7	10	4	5	10	5	41
UO	5	10	3	5	10	5	38
PU	5	10	10	5	9	10	49
XJTU-SY	8	10	8	5	10	10	51
KAU	3	10	8	5	10	5	41

[1] Appropriate Data Amount
[2] Free-of-Error
[3] Consistent Representation
[4] Ease of Manipulation

The following observations are drawn based on the scoring provided for these different categories:

1. Appropriate Data Amount - There is no generalized consensus on the correctly needed data relative to a healthy and failed component in condition monitoring literature. The lack of consensus is further reflected in the datasets generated for bearings, where there are some datasets, such as the CWRU, with very little amounts of data but reflect multiple different failure modes, and the IMS dataset, which has abundant amounts of data but reflects the imbalance seen in manufacturing applications in industry. The other dataset falls anywhere along the spectrum with respect to data amount versus failure classes identified, increasing the confusion in a standardized metric. Hence, the metric in assessing the failure modes, types, operating conditions, and stages, was meant to capture the versatility of application for these datasets based on the documentation rather than the direct amount of data.

2. Completeness - In this application, the data were counted to ensure that all computed data are present as expected for these datasets. In characterizing these datasets, a quantifiable metric related to capturing missing data has not been presented. Hence a ratio was used to assess the number of files and data points based on the documentation. The deviations were noted and checked with the documentation to see if they were present in writing. The metric is not a requirement; however, having missing data in the data generation could lead to doubts and confusion related to further analysis. Hence, the datasets with missing data outside a feasible limit, in this case, 2% per file or collection, were penalized.

3. Believability/ Free-of-Error - There are a number of different metrics to consider when analyzing failed bearings. ISO 15243 [6] provides a comprehensive bearing "investigation" pipeline for analysis when they fail. Including these metrics for bearing analysis should be the standard. There are numerous analysis methods for bearings; however, they generally fall into two categories either time- or frequency-based methods. Both techniques should be the standard in assessing bearing damage criteria. A similar control is placed in material failure studies where damage is measured relative to the number of loading cycles and conditions subjected to the material.

4. Consistent Representation/ Ease of Manipulation - Consistent representation is a standard metric to ensure that the data are presented in the manner described in the documentation. Deviations about the structure lead to doubts related to whether the data reflects the analysis described in the supporting documentation. Ease of Manipulation considers the different methods one can use for the generated data for analysis. Each of the datasets reflects diagnosis categories in comparing healthy versus failed bearings. However, further prediction on the bearing health timeline requires data capture with respect to a metric, such as damage and time.

The scores are not meant to state whether the individual bearing datasets are useful in one category over another. It is more so to demonstrate how is easy it is to interpret the dataset, utilized the gathered information, and apply any general knowledge to a similar situation. That is the purpose of generating these datasets in the long run. The scores provide an easier understanding og which dataset is easier to interpret and apply to another bearing condition monitoring application. With these categories properly, standardized metrics are needed to justify the datasets

understandability, and guiding principles are needed to link these analyzed dataset characteristics into the data generation process.

4.1.1.4 Data Generation Principles

In Section 4.1.1.3, the bearing datasets were originally analyzed with respect to prior condition monitoring literature characteristics. In doing so, the datasets were split into different levels of understandability and versatility based on the associated metrics with each dataset. Six guiding principles are proposed for future dataset generation to better capture these dataset characteristics. These principles differ from prior work, such as Hagmeyer *et al.* [416], which characterize the dataset scenario for generating data. Instead of focusing on a case-by-case scenario, these principles should be captured in every type of condition monitoring dataset generation. The six principles are: Purpose, Testing Methodology, Physical Training Data, Verification & Validation, Data Labeling/Documentation, and Data Transferability. Each principle is linked back to one or more of the dataset characteristics assessed in the last section, along with possible links to the other criteria listed in Table C.2 in the appendix.

Purpose. Each condition-monitoring dataset should have an identifiable mission statement, motivation, or reason for generating the data. The motivation is criteria listed by Hagmeyer *et al.* in their dataset characteristic list. There are formulaic methods for designing mission statements with nine categories describing most: (1) Customers, (2) Products or Services, (3) Markets, (4) Technology, (5) Concern for Survival, (6) Philosophy, (7) Self-concept, (8) Concern for public image, and (9) Concern for employees [435]. The criteria presented are more tailored to businesses; however, the right mission statement has been found to promote further ownership and accountability with the product [436]. The tailored criteria for condition monitoring applications, keeping similar definitions, could consider the following criteria in Table 4.8 in forming a purpose or motivation for generating condition monitoring datasets. The guiding questions for forming the condition monitoring mission statement help start the formulation of the other data generation principles. This guiding principle serves as the basis for the appropriate data amount. The basis of the motivation provides the user with the basis for how much use they can expect from the dataset in terms of what data is available, the type, and the applications it considers.

Testing Methodology. A firm understanding of the testing methodology is needed for application in the real world to verify the damage methods and allow for the recreation of any data generation.

Table 4.8: Mission Statement Criteria

Original Criteria	Condition Monitoring Criteria	Guiding Questions for CM Criteria
Customers	Component	What components are tested?
Products	Failure Mode Data	What failure modes are considered?
Markets	Applications	What applications are covered?
Technology	Data Types	What types of data are included?
Concern for Survival	Data Trends	How does the data change over time?
Philosophy	Testing Process	How are the data generated?
Self-Concept	Data Uniqueness	How are the data unique?
Concern for Public Image	Data Dissemination	How is the data available for use?
Concern for Employees	Data Validity	How are the data verified for use?

Bearings are precision systems and are susceptible to multiple failure modes that are captured in different forms. These different failure modes could manifest in different signals captured during the data analysis. Hence, the accurate detailing of the test methodology links changes in the data with changes in the actual system. However, it also helps provide the users with a level of understanding with respect to the likelihood of appearing in a real system. For example, drilling holes in the bearing raceway is one damage method to generate bearing fault data in bearing condition monitoring. However, the likelihood of a circular fault through the bearing raceway is minimal. Hence, the applicability of that data outside of the research domain is small in comparison to data generated in a run-to-failure rig. Furthermore, the testing methodology organizes the data based on the test, operating conditions, equipment characteristics, and damage applied. Hence, this guiding principle considers the Completeness and Consistent Representation dimensions. The testing methodology provides a basis for verifying the correct amount of data in a valid format.

Physical Training Data. The primary data pool for offline testing should be physical data collected from an experimental system. There are benefits to using synthetic data to grow datasets and increase the likelihood of failure detection [437]. However, the basis for synthetic data requires experimental knowledge to verify the expected failure signatures. Hence, generating physical data can verify the expected failure characteristics in a signal for specific failure modes. The physical data generated could then serve as a basis for synthetic data generation techniques. Similar applications can be found in data augmentation applications. For example, Alqudah *et al.* [438] compares the data augmentation and synthetic data techniques generated of a real dataset for classifying semiconductor defects. Similar applications can occur using Generative Adversarial Networks (GAN) for

synthetic data generation in bearings based on accurate data [366]. The basis for these additional computational techniques relies on a physical source of training data. The dimensions associated with this principle are related to the Free-of-Error dimension.

Verification & Validation. Every condition-monitoring dataset should have verification and validation metrics for the generated data. Verification methods represent quantifiable differences in the data generated, such as statistical changes in time-series data and peaks increasing in the frequency data. Physics-based models could be incorporated into the verification to ensure the correct data output for different failure modes. The validation processes consider the anomalies that happen during the verification process. For the data generation process, the detected faults could assess sensor damage, equipment faults unrelated to testing, or noise from surrounding equipment. The verification process entails procedures to ensure the correct amount of data is included in the dataset based on the test methodology descriptions. This guiding principle corresponds to the Free-of-Error, Ease of Manipulation, and Completeness dimension.

Data Labeling/ Documentation. This principle encompasses the four prior guiding principles into a single document for ease of data dissemination and use in different applications. The complete documentation includes the reasons for creating the dataset, the methods for creating the dataset, the data itself, and the verification and validation procedures used to ensure the data represent the expected scenarios. The data labeling should be conducted in a manner that uses terminology related to the equipment failure modes. For example, defects in bearings should be labeled based on damage related to the bearing failure mode documentation in ISO 15243 [6]. The document provides additional terminology to describe bearing faults, such as contamination, thermal, and chemical damage. This guiding principle corresponds more so to the Appropriate Data Amount and the Ease of Manipulation dimensions. This guiding principle provides a means to standardize and document how data are disseminated after the data generation process.

Data Transferability. The final guiding principle considers the methods to transfer data between different systems. Data transferability describes the methods used to move data, information, and knowledge between different systems. For the knowledge aspect, it is possible to use techniques, such as transfer learning, to apply the knowledge gained from the generated data to similar systems. For the information process, expected limits learned from the generated could be scaled and applied to other systems. For the data transfer between systems, scaling laws are needed to prove a relationship between one system's generated data and an existing system's generated data. One example is the use

90

of similitude in structures to verify the data, information, and knowledge between scaled and actual structures [439]. In condition monitoring systems, these physics models can be complex depending on the application and scenario; however, physics-informed methods could provide approximations for physical attributes that require computationally intense calculations. This guiding principle is considered in the Ease of Manipulation dimension. Currently, most datasets only consider the methods for conducting diagnosis and prognosis analysis. However, providing information related to the system to ensure the data transferability to other systems helps ensure a one-to-one relation between similar systems.

4.1.2 Design and Condition Monitoring Methodologies

Two different areas were considered for the initial methodology formulation: design and condition monitoring methods. For the surrogate system to act as a counterpart to the real system, it must resemble the real-world physical system. However, repeated failure of one-to-one systems would increase the overall cost of the CMDTr implementation. A scaled version would be needed to integrate the surrogate system into the condition monitoring system. However, the system design needs to consider the real-world condition monitoring system and how that is implemented alongside the designed system. The extracted content from these areas helps formulate a process implementable with the design of new systems alongside existing production systems. In each of the following sections (Sections 4.1.2.1 and 4.1.2.2), an overview is provided with respect to a design process and condition monitoring process widely accepted: the *Engineering Design A Systematic Approach* by Pahl and Beitz *et al.* and ISO 17539, a condition monitoring implementation process. From these systems, an ensuing failure methodolgy is created to act in coordination with both methods to form a surrogate system.

4.1.2.1 Design Process

For the design of the surrogate system in the *Purposeful Failure Method*, the design process adopted comes from the Pahl and Beitz Method, termed *Engineering Design A Systematic Approach* [4]. Pahl *et al.* splits the design process into four areas: Planning and Task Clarification, Conceptual Design, Embodiment Design, and Detail Design. During the Planning and Task Clarification stage, the design team begins with a task/problem to solve. The culmination of the ideas related to the task becomes the requirement list. With the requirement list, design process moves into the

Conceptual Design phase, where specific problems are identified, and concept variants are formulated. These concept variants are formulated into a principle solution, beginning the Embodiment Design phase. During the Embodiment Design phase, details, such as the material selection and engineering calculations, are added to this principle solution. The solution is refined to eliminate weak spots and errors and prepare for production. The final phase is the Detail Design phase, where the production and operating documents are finalized with detailed drawings for the solution. Figure 4.2 describes the Pahl and Beitz method based on these four stages and the different deliverables for each design step.

During the planning phase, the underlying motivation is revealed for the product. A market analysis is conducted for existing products and whether a need is present during this design. Engineers determine flaws and existing gaps for current products and create a mission statement (product proposal). Based on a review of Fortune 1000 companies, Williams *et al.* [440] found that the top- and bottom-performing companies had a separable difference in their mission statements. Those top-performing companies had clear, identifiable components that accurately conveyed who the company was and what it did. These criteria have been further found to give employees greater motivation in their drive for their work [441]. In design, the mission statement focuses on the anticipated design, such as the purpose of the final design. From the establishment of the mission statement, further clarification is provided, and a requirements list is formulated.

After the initial requirement list formulation occurs, additional requirements may be added as the design process continues. The initial formulation requires a general framework that allows for easy integration of these checklists as design continues [442]. The basis of the requirements list continues with the formulation of conceptual designs. During this phase, function structures and working principles are fleshed out to link the created requirements list with the principle solution. These function structures help manage the changing states that might occur in the product. The working principles determine how to fulfill the functions described in the functional structures. After a selection occurs using design trees and morphological charts, a principle solution is selected. This solution represents the best initial design characteristics.

The embodiment design phase begins with the principle solution. Examples of work that begins with this phase consider material for different components, and engineering calculations to verify parts of the design. During this time, actual prototypes may be designed for certain functions to verify how they would work. The prototypes would serve as a scaled-down representation
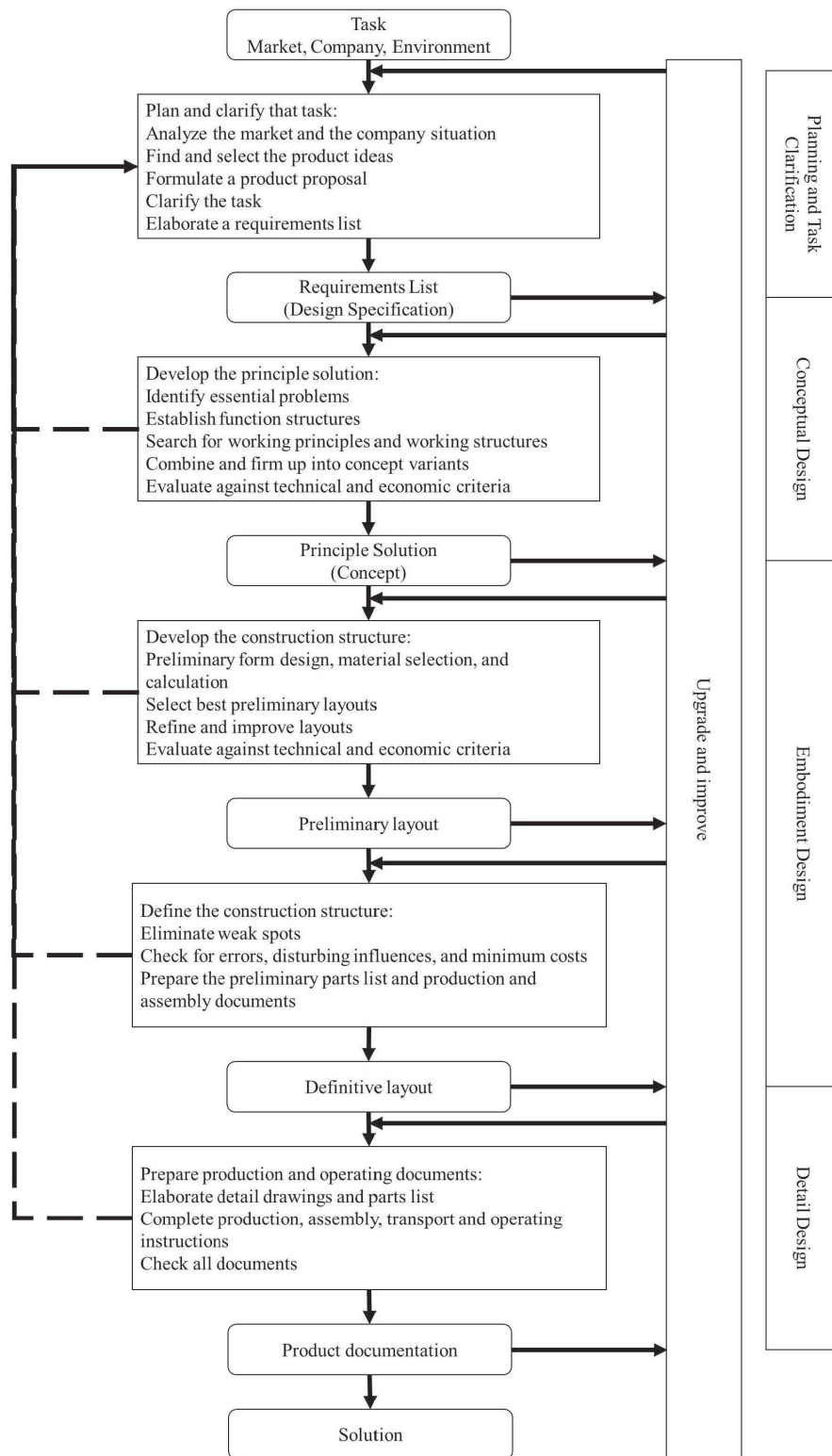
Figure 4.2: The Pahl and Beitz method as described in [4]

using dimensional analysis to minimize the cost and maximize the information needed for these designs. With technologies such as 3D printing, it is quicker to make functional representations and verification during the design phase [443]. As the prototyping occurs, the design is updated with the new information and checked for potential errors using Failure Mode and Effects Analysis (FMEA) [444] and Fault Tree Analysis (FTA) [445]. After meeting all of the checks related to the failure analysis method and passing the verification, the product documentation is formulated and production begins.

4.1.2.2 Condition Monitoring Implementation: ISO 17539

As mentioned previously, the four main foci with condition monitoring systems involve data acquisition, processing, analysis, and maintenance decision-making [71]. However, these are each very broad topics, and if the proper design principles are not incorporated into these measures, then the condition monitoring system is rendered ineffective. ISO 17359 provides the general procedures for implementing a condition monitoring system [5]. The overview steps are broken down into A cost-benefit analysis for condition monitoring, an equipment audit, a reliability and criticality audit, a maintenance strategy selection, a monitoring method selection, data acquisition and analysis, maintenance action determination, and review. These overview steps could be grouped into four categories: the condition monitoring system motivation, the system preparation, the active monitoring state, and the decision-making. Figure 4.3 provides a general procedure outline and how the overview steps are grouped into their respective categories.

An argument is made to adopt condition monitoring during the condition monitoring motivation phase. As discussed in Section 2.2 (Cost of Unexpected Downtime), the primary driving factor for adopting condition-based monitoring is to reduce the cost of equipment downtime. The cost analysis considers area covered in Section 2.2, such as the repair cost, labor costs, environmental impact, and lost production. For example, in the field of wind turbines [446] and rail way costs [447], the benefits of equipment condition monitoring can significantly reduce the capital costs associated with their downtime. However, there are additional costs associated with condition monitoring implementation. Yang *et al.* [448] found that condition monitoring increased the diagnosis and prognosis power for equipment health predictions but condition monitoring was much more expensive over traditional Supervisory Control and Data Acquisition (SCADA) systems. These systems are not mutually exclusive, but there are tradeoffs to consider between using SCADA over
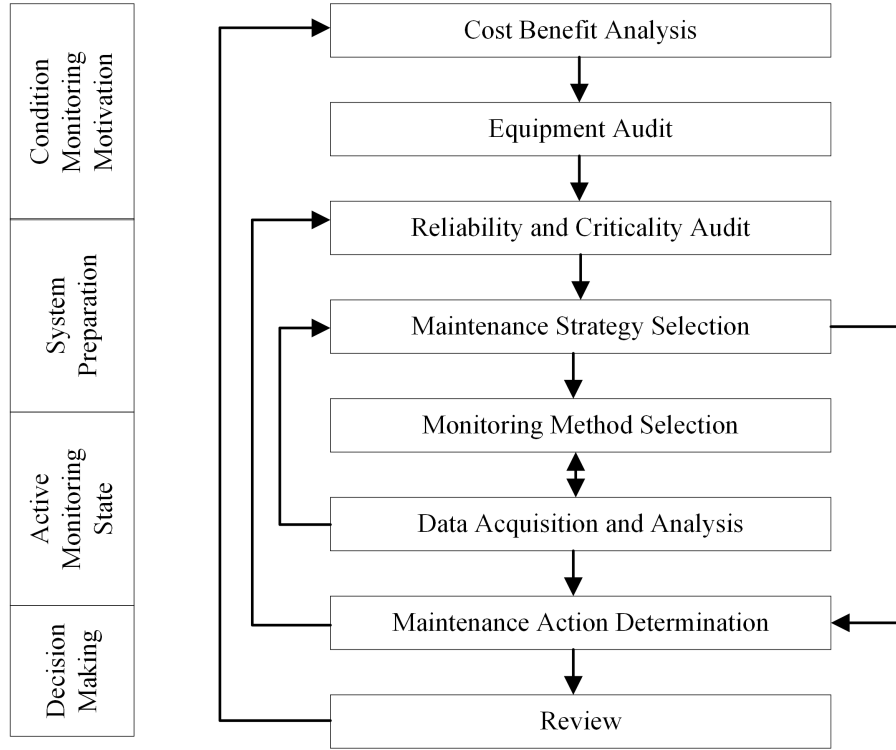
Figure 4.3: General procedure for ISO 17359 [5]

CM systems. These are factored into the equipment audits and reliability analysis to determine what are the cost with the current maintenance practices in place. As found by Adu-Amankwa *et al.* [68], predictive maintenance provided a greater cost reduction than preventive and corrective maintenance; however, that is not a definite fact in every scenario and is shy a proper motivation is needed before implementing condition monitoring.

There is some overlap between the reliability and criticality audit stages for the system preparation phase of condition monitoring implementation. At this juncture, failure analysis tools are used to determine what areas of the equipment are monitored. The FMEA is considered the standard in most design methods and in the manufacturing industry [449]. However, other failure tools, such as the fault tree analysis and root cause analysis methods (*i.e.,* cause-and-effect diagrams) [450] are also employed to map equipment failures to critical functions and provide a rating. A maintenance strategy is eventually chosen, after equipment failure modes and criticality are determined and combined with the prior motivations surrounding cost. If the method is not a CM-based maintenance strategy, then other methods are put in place. A monitoring method is

needed to determine the key performance metrics if a CM-based strategy is used. For industrial practices, various data acquisition methods have been tested and implemented in each scenario. For example, in wind turbines, Liu *et al.* lists vibration, acoustic emission, lubricant, power quality, microscope, and temperature as methods to acquire and implement condition monitoring [388]. The monitoring method should have associated standards related to their implementation depending on another of different standard organizations [451].

The active monitoring stage is when data related to the equipment conditions is collected to start making decisions related to the overall equipment's health. After the establishment of the equipment's normal routine or "baseline", different diagnostic and prognostic controls are determined to asses the degrading state of the equipment. These controls can vary based on the parameters collected from the equipment and the underlying condition. For example, in vibration, there are certain frequencies or vibration limits [389] that can determine when the equipment is about to fail. Additional information can come from equipment manufacturers, such as temperature limits from SKF [94]. However, these standardized limits vary based on the equipment size and configuration. When one of these limits is tripped, though, the decision-making phase for the condition monitoring system begins. During the decision-making phase, a schedule is determined to enact a repair that involves taking the equipment offline. Depending on the CM system, the repair schedule could occur automatically or through the use of a human.

4.1.2.3 Synthesized High-level Methodology Relationships

Generally, a similar process is evident from the design and condition monitoring methods. Both methods begin with the underlying motivation and analysis to determine if there is enough justification for their implementation. Whereas the design method begins with a requirement list to motivate the initial concepts for the product, the condition monitoring system considers the different monitoring methods that enable the maintenance strategy. In the design process, a final concept is chosen with design calculations and considerations to ensure the working order of the final product. For the condition monitoring process, the monitoring method is implemented and tuned to the equipment for the condition monitoring system. During this period, diagnosis and prognosis criteria are implemented and updated as the system collects data. For the condition monitoring process, the final system provides decision-making power to the manufacturers to perform maintenance. For the design process, a final product is designed that fills a niche market. Hence, to integrate PFailM, the

process considers the inputs from the failure analysis study during design and condition monitoring to formulate the damage generation for the component. Data from the data generation is collected using the monitoring methods considered and selected as part of the design and condition monitoring process. The data are related back to the original system utilizing equipment knowledge, such as the geometric, kinematic, and dynamic constraints expected from the surrogate triplet to their related quantities in the physical triplet.

4.1.2.4 Modeling Approaches

The basis for the PFailM approach is centered around using a surrogate system to create an offline environment, as mentioned earlier in Chapter 3. Hence, a surrogate model based on known physics phenomena is employed to consider the another using a Mass-Spring-Damper (MSD) system. The MSD system represents a means of relating data between bearings of different sizes based on understanding the known physical transformations between a source system (surrogate triplet) and a target system (physical triplet). However, other modeling approaches make it possible to establish the relationship between the physical and surrogate triplet when relating data. The different approaches are broadly formed through either numerical or analytical modeling approaches. Additionally, prototyping modeling approaches could enforce the relationships between the surrogate and physical triplet.

Numerical models consider a large number of mathematical equations built around known physical quantities to approximate a solution to a physical system. Von Neumann *et al.* [455] are considered to have the first work in the modern approach for numerical modeling when investigating the errors associated with inverting large matrices. Matrix manipulation is a fundamental part of numerical modeling due to the large number of equations associated with each physical system. Numerical modeling has been implemented for bearing analysis due to known geometric and material constraints to form dynamic parameters [112, 143, 336] through finite element analysis. The vibration formulation can consider nonlinear and linear tendencies for these parameters based on known geometric, kinematic, or dynamic constraints related to the physical system. However, the numerical solutions' performances are based on how well the parameter approximations are conducted and the nodal approximation associated with each modeled object. As the number of nodes increases, the number of equations and parameters increases exponentially, increasing the necessary computation power and time to reach a solution.

Analytical models consider a set of mathematical equations based on specific parameter relationships to determine a solution as it changes concerning time. The type of analytical models used for bearings depends on the desired approximation. For example, in crack fatigue, fracture mechanics are employed to estimate the growth length of a crack using the Paris crack growth model [340, 341, 342]. MSD models [114, 337] are another form of analytical modeling used to approximate bearing vibration based on a known set of equations associated with different physical parameters associated with the system. The physical approximations determine the number of equations needed for the solution, similar to the format of the numerical solution. However, the analytical solution returns an exact response based on the approximations and equations formulated for the analysis. A disadvantage of analytic approaches is that these closed-formed models may require a deeper understanding of the system that may not always be accessible. In this case, the model formulation cannot occur in a closed-form and more numerical approximations are required.

The prior methods consider more so data generation through computational simulation and these are further discussed later for the bearing physics modeling in Section 5.1.1.1. Other forms testing to either a similar size and configuration to the end product expenditures faced by small, high-risk projects in the early 2000s [452] to increase cost savings. The prototype process allows the simulation of final products using scaled models to reduce the cost of testing full-sized objects. For example, Rothhaar *et al.* [453] demonstrated the approach for creating a prototype environment to simulate the change from hovering to wing-born flight on a scaled model. In addition to demonstrating and validating the design of the prototype, the test bed created, as a result, enables successive model testing as design considerations change for the prototype model. However, for the prototype to relate to a real system, the scaled system demonstrates the need for established relationships in the physical quantities between the scaled solution and the final solution.

In this approach of the surrogate triplet, surrogate models are chosen due to their generalized approximation of a system versus the need for high-fidelity quantities found with numerical, physical, and prototyping applications. Surrogate models are formed as mathematical models to determine an outcome that is not easily measured or approximated. Despite the numerous models and methods for approximating bearing vibration, the numerical methods and analytical models are heavily dependent on the parameters approximated. For prototyping, the approximation of the physical parameters can vary greatly based on the physical relationships between the scaled model and the actual application. Hence, the surrogate triplet takes elements of each to create a

surrogate system to approximate the data that could come from a physical counterpart. It should be noted that the physics parameters are still employed based on known constraints related to the system, as overgeneralization does increase the risk of no relation between the different systems. The relationship of the generated data from the surrogate to the physical triplet is then analyzed using a generalized analytical model representing a bearing system in either environment. The data relationship between the physical and surrogate triplet is discussed more in Chapter 5, whereas the system design for the surrogate triplet as it relates to the physical triplet for bearings is discussed in Section 4.2.

4.1.3   Purposeful Failure Method

The PFailM was designed to compliment the previous design formulations drawn by Pahl and Beitz *et al.* [4] and ISO 17359 [5]. The PFailM aims to generate condition monitoring data for generalized applications. Using this method, a surrogate system is designed to resemble the original equipment, whereby failure data is generated to resemble the failure modes anticipated. The system design considers the intrinsic knowledge to build the system and then uses these methods in the transfer back to the original system the surrogate was designed to represent. Some of the method and methodology characteristics are similar to Function Failure Design [454]. Using Function Failure Design, the functions and failures are linked together to improve the original concept. The concept generator phase from Pahl and Beitz *et al.* is modified to focus the concept design on addressing these identified failure modes with their corresponding functions. Fault propagation analysis is another complementary technique that serves to map how faults progress through complex systems [455, 456]. The analysis begins with identifying a failure in a particular function and how it affects subsequent systems. The formulation of these failure modes and process allows for the automatic generation of failure scenarios, such as the work conducted by Irshad *et al.* [457]. They worked to model early human failures in complex systems and how they affected equipment life and failure scenarios. The primary difference between PFailM and fault propagation analysis and Function-Failure Design stems from the generation of data specifically for condition monitoring systems. In this manner, the method purpose is to help augment the training of CM systems for their integration into production equipment by providing instances of failure data before they begin to occur.

The initial formulation of the Purposeful Failure Methodology is meant to be slotted into the design and condition monitoring methods discussed in Section 4.1.2. Figure 4.4 and Figure 4.5

shows where PFailM is meant to be slotted in with design and condition monitoring methods. PfailM is meant to be integrated inside of the definitive layout portion of the Pahl and Beitz *et al.* method. During the time of the definitive layout, the prototype and validation portion could incorporate data generation for differing failure modes when implementing the purposeful failure method. For the condition monitoring implementation, PFailM is implemented alongside the monitoring method selection and data acquisition selection and implementation. During these steps, it is possible that an offline model could be leveraged to serve as a validation or expectation for how the monitoring method performs as equipment degrades.
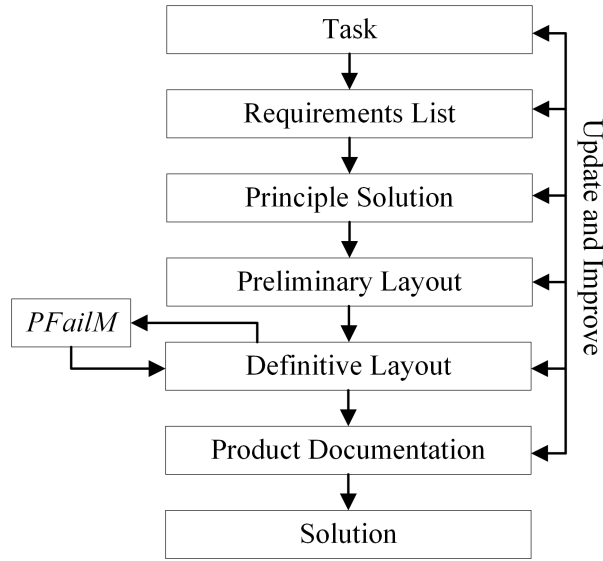
Figure 4.4: The integration of PFailM with Figure 4.2

In general, though, PFailM contains six major steps to cover the design, implementation and generation of a surrogate system to resemble a real-world manufacturing system. The steps consider the data criteria, design process, and condition monitoring implementations to plan out the methodology for standardizing data generation. These steps are: Identify functions and failure modes, Select damage implementation, Determine damage propagation, Select data acquisition methods, Design a surrogate system and generate data, and Verify and transfer data to the real-world system. Figure 4.6 shows the abreviated steps with corresponding sub-steps below. Sections 4.1.3.1 through 4.1.3.6 consider the additional reasoning behind each step in the method. During each step, parallels are drawn to the aforementioned areas of data criteria, design, and condition monitoring implementations in both their similarities and where they diverge.
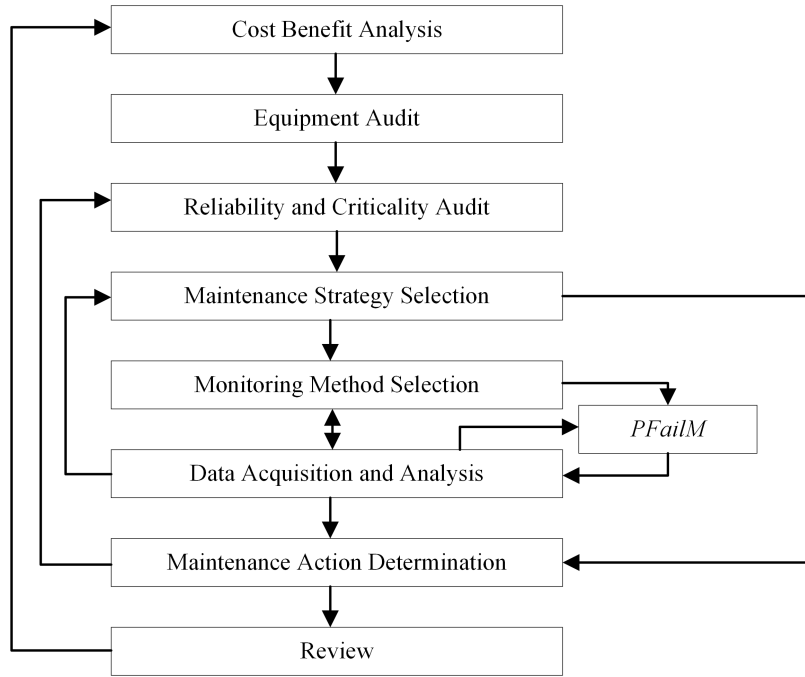
Figure 4.5: The integration of PFailM with Figure 4.3

4.1.3.1 Identify the Functions and Failure Modes

Step Purpose: To identify the functions most critical to the equipment and production operation whereby a failure could cripple the greater functionality over the plant environment. Determine the failure modes that would affect the critical equipment functions through there different methods of causes and effects.

The function and failure mode identification is the first step in PfailM primarily because it is a common enough starting point between the data criteria, design process, and condition monitoring implementation. In this step, the motivation and different failure modes are determined while adding contextual knowledge for the overall system. In some condition monitoring systems, the identified functions are easily determined; however, in complex systems the link between function and failure modes can be lost. Hence, the reason for other methods such as the Function-Failure Design method [454]. The initial formulation for the methodology was to consider components, such as bearings shown in Figure 4.7. However, the method can extend to how robots may fail over time and different processes as well.

The equipment functions should be readily available based on the equipment documentation or knowledge and how they are integrated into the existing process. Depending on the level as shown

```
┌────────────────────────────────────┐
│ Identify:                          │
│ Equipment Analysis                 │
│ Failure Mode Identification        │
└────────────────────────────────────┘
                 │
                 ▼
┌────────────────────────────────────┐
│ Select:                            │
│ Failure Mode Criticality Assessment│
│ Ideation for Recreating Damage     │
└────────────────────────────────────┘
                 │
                 ▼
┌────────────────────────────────────┐
│ Determine:                         │
│ Damage Propagation Study           │
│ Health & Lifetime Indicators       │
└────────────────────────────────────┘
                 │
                 ▼
┌────────────────────────────────────┐
│ Select:                            │
│ Data Acquisition Systems           │
│ Collection Strategy                │
└────────────────────────────────────┘
                 │
                 ▼
┌────────────────────────────────────┐
│ Design and Generate:               │
│ Surrogate System Design            │
│ Data Generation                    │
└────────────────────────────────────┘
                 │
                 ▼
┌────────────────────────────────────┐
│ Verify and Transfer:               │
│ Failure Analysis                   │
│ Data Transferability Assessment    │
└────────────────────────────────────┘
```

Figure 4.6: The Purposeful Failure Methodology

in Equipment 4.7, the function list can stem from one or two at the component level to hundreds at the process level. As an example, a function at the component level may involve translating motion between two subsystems as is done with a bearing. The function of a gear is to increase the force or increase the speed between two subsystems. At the system level, a cobot robotic arm (*i.e.,* a UR10 or Kuka Iiwa) system could consider joints as multiple subsystems, which are in turn made up of multiple different component, such as bearings and gears. There may be an overall function to the robotic arm; however, in response each of those components for the different subsystems and components require consideration as the failure of one could offset the control of the overall system. The level and depth varies depending on the starting equipment level. In truth, the determination of those critical functions is up to the user. Using design processes, these functions are easily
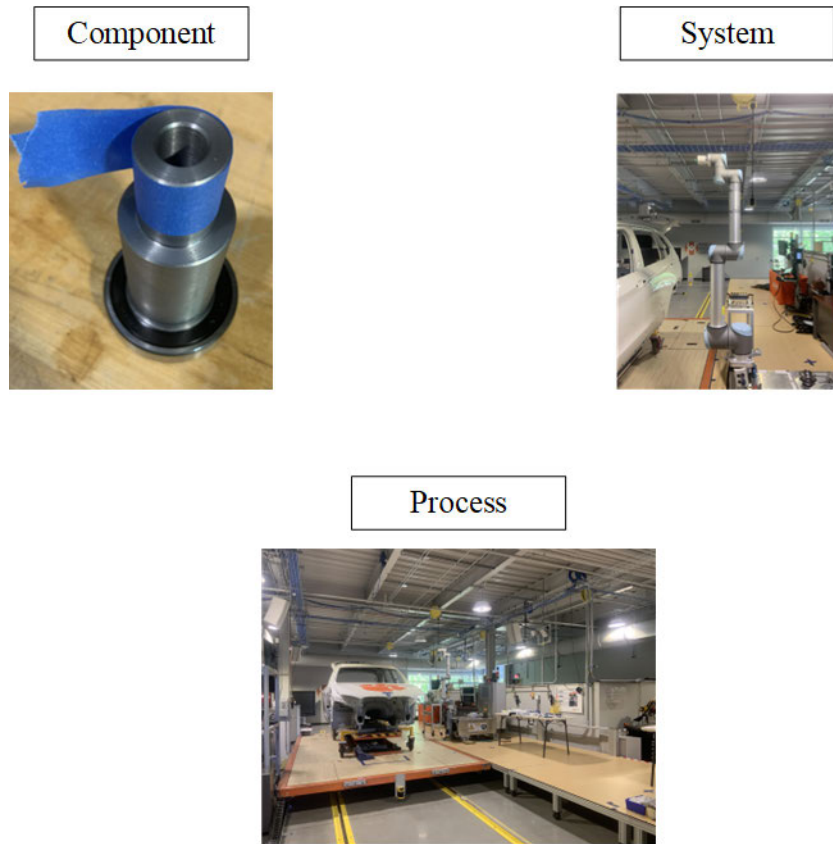
Figure 4.7: The different equipment levels considered under PFailM

disseminated if not provided on hand by any equipment documentation.

After determining the functions, the secondary portion of the step is the failure mode determination. Steps and methods already exist in place to determine the equipment failure methods. The preferred tool to use for this analysis is a Failure Modes and Effects Analysis (FMEA) [444]. Figure 4.8 contains a breakdown of the different failure events as they correlate to identified functions. There are nine columns for an FMEA with a tenth column added to make potential design changes. Each column helps to organize the failure modes and rate them against other failure modes with other equipment functions to determine and address failure modes in a certain order. For PFailM, the FMEA organizes the failure modes clearly to determine which methods to use for recreating failure in the following system. The knowledge contained in the table is necessary to consider in the following steps.

Goldberg *et al.* [458] published a technical report from NASA that provided a list of 15 different methods to consider in system safety reliability and analysis tools, among which an FMEA

Equipment Function	Potential Failure Mode	Potential Failure Effect
What is the Equipment Function?	*In what ways can the step go wrong?*	*What is the failure impact?*
Bearing – Rotational Motion	Fatigue Damage	Bearing Destruction, Loss of Rotation and Support

(a)

Severity	Failure Causes	Occurence
How severe is the impact?	*What causes the failure to occur?*	*How frequently does the damage occur?*
1 – 10 (10 is a catastrophic event)	Inadequate lubrication, long life, contamination, damaged installation	1 – 10 (10 is a frequent event)

(b)

Process Controls	Detection	Risk Priority Number (RPN)
What are the existing controls?	*How likely is detection?*	*How frequently does the damage occur?*
Periodic vibration measurements, speed monitoring	1 – 10 (10 means a low probability of detection)	Severity * Occurrence * Detection = RPN

(c)

Figure 4.8: Example FMEA record showing bearing fatigue damage mode. In general, these columns are represented all in one row. (a) contains the Equipment Function, Failure Mode, and Failure Effects, (b) contains the Severity, Failure Causes, and Occurrences, and (c) contains the process controls, Detection, and Risk Priority Numbers.

was listed as one of the methods. Other failure analysis methods considered FTA, Probabilistic Design analysis, and Cause-Consequence analysis. These methods are useful for failure mode identification and could substitute or supplement the knowledge gained using an FMEA. Scriboni *et al.* [459] considered the use of FMEA and FTA to identify and track the effectiveness of different failures within a UR robotic arm. They showed that similar failure modes were identified with similar risk probabilities. Song *et al.* [460] used FEA to model potential failure modes within wind turbines and

determine the potential failure effects throughout a complex system. Abdelghany *et al.* [461] used cause-consequence graphs to model reliability in electrical networks. They utilized an existing set of failure modes to consider the different cascading effects throughout the system. The failure analysis tool selected should be able to provide comprehensive knowledge on what causes failure in the end and how it affects the existing system.

Step Outcomes: This step determines an understanding of the equipment functions and failure modes. Through the use of design and failure tools, already widely used in industry, documentation is provided for the motivation in future failure testing and how the testing links back to the physical system from the surrogate system. Depending on the equipment level, the number of functions and failure modes varies and may drive the creation of more than one surrogate system.

4.1.3.2 Select the Damage Implementation

Step Purpose: Determine the critical failure modes from the conducted failure analysis in Step 1. Consider the different failure effects and select different methods of recreation to induce defects of similar scale in the bearing system.

Damage implementation methods come from the original failure analysis conducted in Step 1. The failure analysis tool determines the failure causes and effects. A ranking is established as a result of the assessment. A criteria is then selected to determine what failure modes are recreated for the system. The criteria for determining the number of failure modes can vary. It could be determined based on the Risk Priority Number (RPN) from the FMEA using a threshold or percentage or based on a subscore from two categories (Severity, Occurrence, or Detection). Depending on the failure causes different failure modes are recreated using different means to create life-like damage. For a component, such as a gear, researchers have chipped and damage teeth to elicit a failing gear frequency [462, 463]. In both instances, the researchers have listed the damage method and additional factors of size and location to relate the damage to the entire system. The method chosen should be repeatable and ensure minimal disturbance to the rest of the system. System-level equipment, such as robotics, could consider the use of a limiter to induce backlash on a robot joint [464]. The subsequent deviation could simulate an anomalous condition affecting the overall system. Run-to-failure testing is another option, where a system is run outside of its normally expected failure modes and then stopped when a condition is met. This testing is commonly used for bearings[408, 409]; however, if configured incorrectly at start-up, it can lead to widely varying

data responses. Incidentally, using purposeful failure methods, these methods must be controlled with tools to enable the best recreation of damage for data generation method. Figure 4.9 shows a bearing damaged in a late-stage defect. It seems that the damage induced is meant to resemble a late-stage defect due to the surface area of damage; however, the crisscrossing lines seem to indicate an uncontrollable nature during the recreation.

Figure 4.9: Example of bearing damage through purposeful failure means.

Step Outcomes: The method for damaging a system is selected and the means are provided for how to recreate the cause and subsequent damage effects to replicate physical damage.

4.1.3.3 Determine the Damage Propagation

Step Purpose: Determine metrics to determine the scale of damage as it occurs during test. Determine health indicators for potential changes during data collection and generation.

Damage propagation is defined as the steps to increase damage during testing. In the methodology framework, there are two types of damage possible in the system: "Run-to-Failure" and "Artificially Generated Failure Data". Run-to-failure testing occurs when a component or system runs to a predefined failure point after a certain length of time. The predetermined failure point indicates some kind of condition that has occurred on the equipment that is induced by the starting conditions. In the bearing dataset generation from the IMS, FEMTO, XJTU-SY bearing datasets, the bearings are overloaded in terms of there applied load and in relation to there fatigue load limit [408, 409, 414]. For materials, there are cyclical cycles that flex the material until it reaches its breaking point. Damage can also be measured after a certain number of cycle have passed in the system and a full breakage has not occurred. The other possible method is to use accelerated failure testing. At certain points in a system predetermined amounts of damage or an

adverse condition to speed up degradation relative to the overall system lifetime [465]. In bearings, this could be the application of small defects to the bearing raceways and rolling elements. The CWRU, PU, UO, MFPT, and KAU [410, 413, 412, 411, 415] used seeded bearing failure data by marking the bearing surface using abrasive tools. By using the periodic damage, failure is sped up; however considerations are needed to ensure that the bearing damage retains the like like expectation as damage increases.

Health indicators provide different measures of equipment life as the damage increases in the system. In battery testing, the capacitance determines what the life stage for the equipment [466]. As a battery degrades, the holding charge reduces over time. The sensory data indication from the testing can provide the indication of degrading health. In the FEMTO bearing dataset [409], the stopping condition was based on the bearing vibration amplitude. The stopping condition for the IMS bearing data [408] was the amount of material removed during testing and was deposited in the bearing oil. Hence, these health indicators should correlate to the equipment degradation to convey how damage changes over time.

Step Outcomes: Metrics are assigned to characterize the change in equipment health as it relates to the damage implementation method. Assigned at each stage of damage are health indicators to signify the changing equipment state and provide a stopping condition for any induced damage.

4.1.3.4 Select the Data Acquisition Systems

Step Purpose: Select the data acquisition system for the surrogate system based on the original equipment design.

The data acquisition methods are a necessary step to ensure that the data collected in the surrogate system replicates the configuration in the real world system. For equipment condition monitoring, vibration, temperature, acoustic emission, lubrication provide measurable data to changes in the equipment condition. However, each of these methods require tuning depending on the application. For example, any vibration system must ensure that the Nyquist frequency/criterion is met. The Nyquist frequency is the minimum sampling frequency needed to measure a signal and prevent anti-aliasing in the signal. For acoustic measures, directional microphones filter out background noise in non-measured zones. For more detail on these different data acquisition methods, the reader is directed back to Section 2.4 in Chapter 2.

Th other consideration from this step for the user is how to combine additional data ac-

quisition methods to increase confidence in any prediction. Sensor fusion is a common application for consideration due to the lower cost associated with computational resources and the widespread application of multiple sensors. Sensor fusion is the process of combining sensor data from multiple different disparate sources to increase the amount of available information [467]. The combination of different data sources allows for greater accuracy in reaching the correct conclusion, where in some situations, the signal is confused when using only one sensor. This step provides the user to bring in additional sensors to test new data streams. These new data stream could then be added to real world system depending on their impact to the overall system.

Step Outcomes: The data acquisition systems are documented in relation to the real world system and deployed to the surrogate system and integrated to the test plan.

4.1.3.5 Design the Surrogate System and Generate Data

Step Purpose: Design and build the surrogate system using design relationships of similitude to relate the experimental system to the real world system. Generate the data by combining the previous four steps into an experimental plan with the corresponding failure scenarios.

During this step, the experimental procedure is formulated based on the prior four steps, the surrogate system is designed, and the data generation begins. The experimental procedure begins with a description of original equipment and why the equipment requires the generated data (*i.e.,* verification and validation of design, missing training data *etc.*). The failure modes are listed in order of priority with the corresponding damage implementation methods and the metrics for damage propagation. At this point, the formulaic procedure describes how to create the associated simulated damage in the surrogate system.

After formulating the experimental procedure though, the secondary consideration involves the preparation of the surrogate system. One manner to view the surrogate system is as a prototype to the real world application. Building scaled prototypes is a common enough occurrence within design to validate the method and performance for how the system performs in the actual environment. A common application in the aircraft industry is to use additive manufacturing to build scaled prototypes for aircraft parts and provide verification and validation of the components using blockchain [468]. He *et al.* [469] showed how digital twins could also use physical prototypes in their design phase to demonstrate different control structures as they move within a system. Furthermore Coutinho *et al.* [443] provided a review of prototype concepts for implementing similitude

for designing prototype buildings with real counterparts.

In prior condition monitoring applications, the primary emphasis has been to replicate systems that occur in industry. In PFailM, a further step is taken to ensure a level of design similitude between the real world applications and the surrogate system. Similitude is defined as a system that meets geometric, kinematic, and dynamic relationship goals between the a scaled environment and the real world environment [470]. The early discussion of the similitude theory could be traced back to a *Dialogues Concerning Two New Sciences* by Galileo Galilei [471] and have been used in a variety of engineering applications. Vassalos *et al.* [470] used similitude as a means to simulate how scaled marine structures operate in the real world environment. Another example of the use of similitude is through the application of wind tunnels [472]. Wind tunnels simulate the effects seen on aerospace structures in the real world through a scaled representation. Casaburo *et al.* [439] considered the use of similitude to design and build scaled models of structures to validate how they would act in the real world.

Similitude is capable of being carried out using dimensional analysis. Dimensional analysis is the study of how to relate physical quantities through the identification of their base relationships [473]. The concept was first introduced by John Fourier in 1822. Since then, the common methods for applying dimensional analysis is using Rayleigh's Method through the formalization of the Buckingham Pi Theorem [474]. During the process, the scaling laws are derived to relate the physical quantities of different systems to each other. These applications occur in vibration analysis as a means of verifying the experimental response to see in such systems [475]. Through the use of similitude, it is possible to design a scaled system for the surrogate system for condition monitoring. The underlying derived relationships are utilized to then transfer data from the surrogate system to a real world system.

Step Outcomes: The underlying experimental procedures and systems are designed. Data generation occurs according to the procedures with the experimental system. During the data generation, documentation should occur ensure the proper operating conditions and labels are assigned the data after the data generation.

4.1.3.6 Verify and Transfer the Data

Step Purpose: Verify the presences of the failure criteria in the bearing data. Implement physics-informed and physics-based data to link the surrogate system data back to the original equipment

in operation.

The first portion of this step considers how the generated data are analyzed to verify damage is present in the data generated. For this step, an exploratory data analysis is considered to determine the different changes in the data. Tukey *et al.* [476] provided a number of different techniques in his published work on exploratory data analysis (EDA), such as box-plots, histograms, and bar charts. The analysis could occur with features or raw data to determine overlap between different distributions. For very clear distinctions in the data, the EDA could differentiate the different types of fault conditions from baseline scenarios.

However, physics informed measures are adopted to confirm the potential slight changes that might occur between systems monitored through condition monitoring. The techniques vary between the different condition monitoring applications and data types. For vibration, acoustics, and ultrasonic data, frequency analysis is a common metric to determine the changes in data related to changing states for equipment health. For example in bearings, there are four rotational frequencies that increase in amplitude as damage increases. For gears, there is a tooth mesh frequency that becomes apparent as cracks in the teeth occur. Digital and physical filters may need to be employed to ensure that the correct signals are captured and remove potential noise from the data.

Data transfer occurs after the verification that confirms the data represented contain the necessary failure criteria. The data transformation occurs using either data-driven or physical parameter analysis through the dimensional transformation and scaling laws created for the similitude scale. The combination of the data-driven and physical analysis changes based on the overall system requirements. One method to verify the data transfer is by using data overlap [477]. The overlap will determine what data are transferable and what do not conform to the physical system. Those data may be discarded or further transformed to determine the deviations in the data.

4.1.4   Section Recap: Purposeful Failure Methodology

This section provided the basis of answers to two questions:

i.) Research Question 1.1: How are the existing bearing datasets deficient for training manufacturing condition monitoring systems?

Answer: ✔ Based on prior literature for Pipino *et al.*, Hagmeyer *et al.*, and Lessmeier *et al.*, there are identified criteria for assessing datasets and characterizing their effectiveness for different data

scenarios as was attempted by Hagmeyer *et al.*. However, in terms of condition monitoring data, there has not been a deeper analysis of how well these datasets meet the actual criteria. In the conducted analysis, it was found that the datasets with a greater amount of contextual knowledge performed better against the criteria from Pipino *et al.* in terms of a general analysis to describe data attributes. While some datasets did perform well, there were lacking characteristics related to understanding the data failure criteria, the transference of data between different failure modes, and the labeling of different bearing failure modes. To clarify, while each dataset had healthy and failure data separated, it was difficult to understand the differences in the data and the failure data generation in some respects.

ii.) Research Question 1.2: What methodology criteria are necessary to fill gaps for condition monitoring datasets?

Answer: ✔ By analyzing design and condition monitoring methodologies, a set of criteria were extracted to ensure standardized data generation. The analyzed steps were: Identity functions and failure modes, Select damage implementation, Determine damage propagation, Select data acquisition methods, Design a surrogate system and generate data, and Verify and transfer data to the real-world system. The dimensions identified from Pipino *et al.* are captured in the steps. The Appropriate Data Amount dimension is ensured by determining the expected failure modes, their damage stages, and varying operating conditions. Planning out the data acquisition system within the surrogate system should ensure Consistent Representation and Ease of Manipulation for any data generated. The data's Believability, Free-of-Error, and Completeness dimensions are verified using the data verification and transfer method in the final step. These dimensions are tied to prior condition monitoring criteria to cement their position and ensure data standardization.

4.2   Failure Methodology Design: Bearings

The generated failure methodologies are created as a means to investigate RQ1.3 using the PFailM from Section 4.1:

i.) Research Question 1.3: How much data are needed to train a bearing condition monitoring system for diverse failure modes?

This section investigates and standardizes the different metrics in identifying bearing damage

111

and formulates them into four damage methodologies: Fatigue, Contamination, Lubrication, and Brinelling defects. These failure methodologies are tied to generalized bearing failure modes. Each subsection with Section 4.2 describes a step related to the PFailM implementations for each bearing failure mode.

4.2.1   Identify Functions and Failure Modes: Bearings

Bearings are common applications with very defined functions, as discussed in Section 2.1. The primary focus of this dissertation work is conducted with deep groove ball bearings. From SKF, a major bearing manufacturer, Deep Groove Ball Bearings [478] are versatile and widely used. They can support radial and axial loads with low friction, noise, and vibration. The primary purpose of this bearing type is to provide a transfer of load and motion between two different systems in a range of speed applications. Hence, motors, robots, gearboxes, and CNC equipment use deep groove ball bearings.

As mentioned previously, bearings are a common failure point in rotational equipment. Due to their wide application, there are multiple different causes of failure. SKF identifies four probable causes of failure: Lubrication, fatigue, contamination, and improper events (operating or assembly errors) [7]. A cause of failure from lubrication is listed as either using the wrong kind of lubrication or the incorrect quantity of lubrication. Fatigue failure stems from the bearing reaching the end-of-life stage and has overstressed certain points on the bearing. Contamination causes failures due to different types of particles entering the bearing raceways and interfering with the rolling elements. Finally, the improper events cause improper bearing operation or assembly failures. An improper bearing operation situation corresponds to using a bearing in the wrong design situation, such that it exceeds the load or speed limits of the bearing.

As these are the potential causes of failures in bearings, there are categorized effects of bearing failures documented by ISO 15243 [7]. ISO 15243 describes the different bearing failure modes categorized into six major categories with 14 distinct failure types across the different categories. Figure 4.10 shows the organization for the different categories of failure and their corresponding types. The general meaning for each failure mode is as follows:

i.) Rolling contact fatigue failure occurs from repeated stresses during operation from the rolling elements passing over the surface of the raceways. Subsurface means that the cracks begin
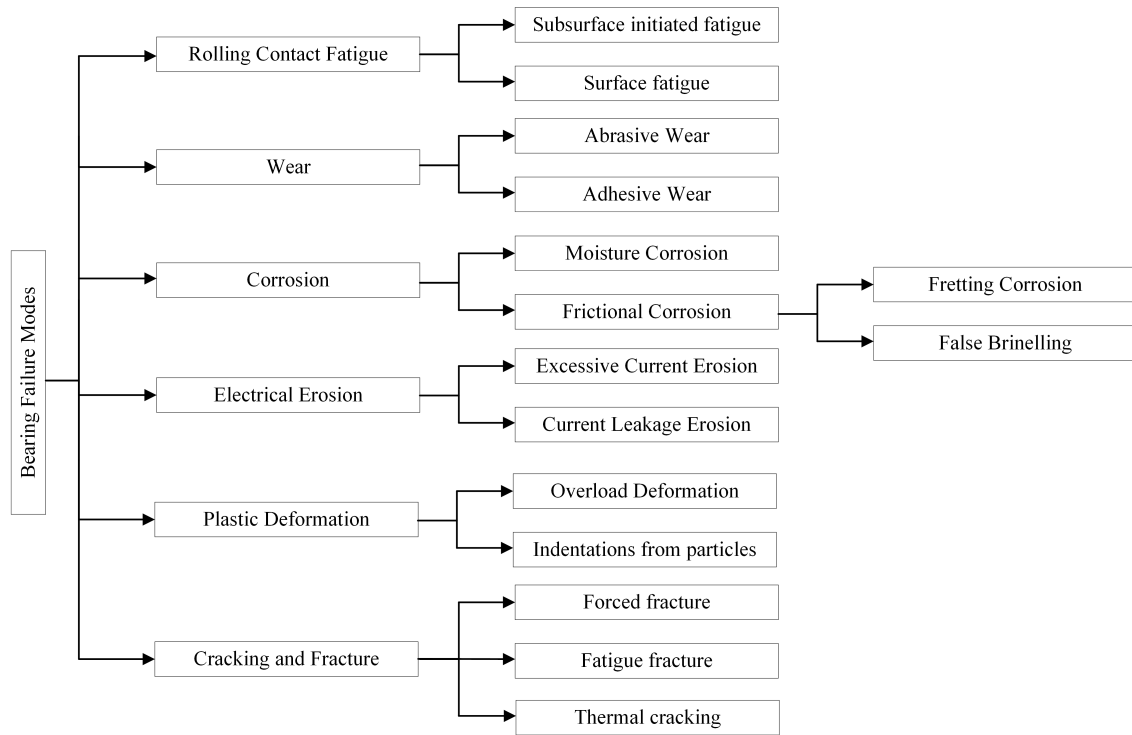
Figure 4.10: Bearing failure modes from ISO 15243 [6]

below the bearing surface, and surface fatigue is caused by asperities on the surface going through plastic deformation.

ii.) Wear describes bearing failures that occur due to surface material removal related to sliding or rolling contact between the raceways and rolling elements. Abrasive wear is caused by the sliding presence of hard particles over the bearing surface. Adhesive wear is caused by the smearing of material between one surface to another surface.

iii.) Corrosion occurs due to a chemical reaction with the bearing surface. Moisture corrosion is caused by the bearing surface coming into contact with some aggressive substance that causes rust to form. Frictional corrosion occurs due to the micromovements between the surfaces, leading to the generation of rust. Frictional occurs from fretting at the fit interfaces and false brinelling from vibration.

iv.) Electrical erosion is the localized microstructural change through an electrical current. Excessive current erosion occurs when an electrical current passes from one bearing ting to the other through the rolling elements and their lubricant films. Current leakage erosion occurs

113

continuously as electrical current passes through the bearing, creating a patterned response on the bearing raceway.

v.) Plastic deformation is caused when the yield strength of the material occurs. An overload deformation happens when the bearing is stationary and is caused by a shock load. Indentations from particles occur when small elements are overrolled between the rings and the rolling elements.

vi.) Cracks occur when the material's ultimate tensile strength is locally exceeded, and fractures occur when the crack occurs through a component section. Forced fracture occurs due to stress concentrations. Fatigue fracture occurs when the fatigue strength limit in bending, tension, or torsion results in cracking. Thermal cracking occurs from frictional heating in the bearing.

The above list considers the general possible failure causes and resulting failure types. However, how these occur in industry depends on the equipment and condition. For example, there is a low probability for an electrical failure to occur in a bearing, not an electrical application. These failure modes are organized into an FMEA, where each failure type is considered a possible failure mode. The FMEA conducted is based on the maintenance documents from SKF [7] and Koyo [479]. These are both maintenance documents that describe bearing failure modes.

Figure 4.11 contains the FMEA breakdown for the general failure modes for bearings. The failure causes and effects primarily come from SKF [7] and are further supplemented by the document from Koyo [479]. The primary reason is that SKF follows the ISO classifications for documenting bearing failures. The failure severity determines whether the bearing damage effects contained fractures (10) versus surface damages (7). These are the two common damage types listed outside of a change in appearance to the bearing elements, as documented in each failure mode. The failure causes were determined based on whether there was a strong correlation to the particular failure mode, as documented by SKF and Koyo. The exception to this rule was for fatigue failures, as these are common regardless of failure causes. Hence, even weak correlations were recorded for these cases. The failure occurrence is determined based on the representation of the four different failure causes as listed by SKF earlier: Lubrication, Fatigue, Contamination, and improper events (operating or assembly errors). The maximum possible score was 10, with the point values broken down as follows: 4 - Lubrication, 3 - Fatigue, 2 - Contamination, and 1 - improper events. These are ranked based on the distribution of failure causes from bearings listed in the SKF maintenance document [7].

Equipment Function	Potential Failure Mode	Potential Failure Effect	Failure Causes	Controls	Sev.	Occ.	Det.	RPN
Bearing Rotational and Load Support	Subsurface Initiated Fatigue	Surface damage due to spalls	Low lubrication Liquids contaminats in Low speed Mounting error Low lubrication viscosity		7	10		70
	Surface Initiated Fatigue	Surface damage due to spalls	Low lubrication Mounting error Solid contaminants in lubricant Mounting error Poor fit High/low speed Poor lubrication quality High/Insufficient load Low lubrication viscosity		7	10		70
	Abrasive Wear	Surface wear on raceway and rolling elemets Band marks on rolling elements Shiny appearance on raceway and rolling elemnt Wear cage	Low lubrication Oxidized lubrication Low lubrication viscosity High load Mounting error Solid contaminants in lubricant		7	7		49
	Adhesive Wear	Discoloration in layers and deposits Ring fracture with discoloration Smear marks Low discoloration Wear Cage Polished surface on raceway and rolling elemets Ring fracture with smearing	External heat Mounting error Insufficient load Low speed Poor fit		10	1		10
	Moisture Corrosion	Discoloration on the bearing surface Corrosion marks Local corrosion Ring fracture with corrosion	Liquids contaminats in lubrication Handling		10	3		30
	Fretting Corrosion	Ring fracture with corrosion Reddish surface on raceway and rolling elements	Poor fit		10	1		10
	False Brinelling	Fluting surface damage from particles on raceway and rolling elemets Standstill vibration marks	Vibration exposure Current passage		7	1		7
	Excessive Current Erosion	Current passage craters	Current passage Current leakage		7	1		7
	Current Leakage Erosion	Fluting surface damage from particles on raceway and rolling elemets	Vibration exposure Current leakage		7	1		7
	Overload Deformation	Impact ring fracture Score marks Surface damage from single point impacts Plastic deformation	Mounting error High load		10	1		10
	Indentations From Particles	Surface damage from particles on raceway and rolling elemets	Liquids contaminats in lubrication High/Insufficient load High speed Low lubrication viscosity		7	7		49
	Forced Fracture	Fracture in one of the bearing rings Ring fracture with discoloration Surface damage from nicks, grooves Impact ring fracture Ring fracture with smearing	Poor fit Material Mounting error High/Insufficient load High speed		10	1		10
	Fatigue Fracture	Ring fracture with discoloration Impact ring fracture Surface damage from nicks, grooves Fracture of cage Ring fracture with smearing	Mounting error High/Insufficient load High/Insufficient load Solid contaminants in lubricant		10	3		30
	Thermal Cracking	Discoloration in layers and deposits Impact ring fracture Ring fracture with discoloration Surface damage from nicks, grooves Ring fracture with smearing	External heat Mounting error High/Insufficient load High speed		10	1		10

Figure 4.11: Generalized FMEA focusing on the bearing severity and occurrence concerning the failure causes and effects from [7]

The detection and potential process controls were not listed as part of this FMEA, primarily due to the number of possible methods for reducing these potential failures. Based on the failure modes, the common control to reduce bearing failure is proper installation, operation, and continuous maintenance to ensure smooth operation. Standards are listed for the proper installation procedures with the bearings from their respective companies. In addition, there are several methods for detecting bearing failure within condition monitoring technologies alone. It was felt that there was no discernible difference in the present scores by adding these values to the RPN calculation.

Based on the RPN calculations, the higher RPN calculations come from those failure modes with possible failure cause categories in general and not necessarily from a higher severity. For example, in the fracture and cracking sections, the failure causes come from improper design and installation. Hence, the lower risk number if there is an assumption that the equipment is professionally maintained. The highest RPN values, outside of fatigue, come from those most affected by damage to the bearing elements caused by a mixture of contamination and lubrication. The fatigue represents the only natural failure mode that could occur in bearings. As a note, it should be considered that these values vary based on the application.

Extracted Information: From this step, the general bearing failure modes are matched up to potential failure causes and effects. In terms of RPN, the top four bearing failures with the highest risk numbers correspond to fatigue cases, abrasive wear, and indentations from particles. These different failure modes provide a basis for determining the damage implementation methods and the damage progression stages.

4.2.2   Select the Damage Implementation: Bearings

The failure mode descriptions best summarize the damage implementation methods for bearings. As mentioned earlier, the primary failure causes for bearings come from lubrication, contamination, fatigue, and operating/assembly errors. The failure effects associated with these different bearing failure modes cause surface damage to the bearing raceways or rolling elements. These can cause fractures and cracks in the bearing elements in certain severe cases. However, before that fracture point, craters, dents, and defects form in the bearing surface to represent progressing damage. For this bearing failure testing, four different bearing damage methods represent the potential failure causes and the subsequent effects. The methods are described in each subsection as follows: Fatigue (Section 4.2.2.1), Lubrication & Contamination (Section 4.2.2.2), and Bearing

Assembly Defect: Brinelling (Section 4.2.2.3). Each of these sections describes the methods for how the damage is applied to the bearings. Section 4.2.3 describes the stages of damage after the initial damage implementation and the stopping conditions associated with the damage progression.

4.2.2.1 Bearing Fatigue

Bearing fatigue is the commonly tested failure mode in research and data generation applications. Each of the eight datasets in the dataset criteria section (Section 4.1.1) investigates this failure mode. The investigations use either "natural" or "artificial" damage to show the growing defect over time. Natural failure testing involves running the bearing to some predetermined point of failure. In the IMS dataset, the end condition was once a significant amount of bearing material was removed from the bearing raceway and rolling elements. In the FEMTO, PU, and XJTU-SY datasets, the stopping condition was determined based on a vibration limit for the system. The other datasets use methods of artificial damage. Artificial damage for fatigue cases involves inducing defects that resemble fatigue damage on the bearing surface. The CWRU dataset used an EDM machine to create their bearing defect sizes. The PU dataset drilled holes and used an engraver to create their defects for their artificial tests. The KAU dataset does contain data from the CWRU dataset but does not clarify how they generated their defects. The MFPT and UO do not say how they generate their defects other than the locations.

A Dremel tool and tabletop CNC were used to test and induce defects on the bearing raceways to replicate the fatigue damage for this testing. Due to the bearing material's hardness, an abrasive tool bit works away at the bearing surface in each instance. The Dremel tool was used initially to determine if enough material would be removed to induce a defect of the expected size. However, there were concerns about controlling the defect size as the defect became larger. The tabletop CNC was a Nomad 3 from Carbide 3D. The spindle ran at 24000 RPM and was powered by a 130 W Brushless DC motor. The tool used to induce the defect was a burring tool designed for grinding applications and performing contouring. The defect sizes were based on the bur tool size. The bearings for this type of testing had polymer cages to allow clear access to the bearing raceway. Figure 4.12 shows the test area with the tool for the fatigue fixture case.

Figure 4.12: CNC Engraving Test area

4.2.2.2 Bearing Contamination & Lubrication

The most common unexpected bearing failure mode stems from poor bearing lubrication. The lubrication inside a bearing is crucial to reducing the metal-on-metal contact that would otherwise occur [480]. A rolling element bearing should reach elastohydrodynamic lubrication if it is well-lubricated. Elastohydrodynamic lubrication occurs due to the rolling motion between the rolling elements and the inner raceway and the resulting high pressure from the elastic deformation [481]. As this contact occurs, the lubricant viscosity increases and maintains a film around the elements to reduce the metal contact. There are several different methods for calculating the fluid film thickness in bearings, such as using the Reynolds Equation and the Etrul-Grubin approach. Both methods are detailed in the review by Lugt *et al.* [481]. The underlying equations are provided in Equation 4.1 for using Reynolds Equation and Equation 4.2 for the Etrul-Grubin approach. h is the fluid film thickness at the specified x- and y-coordinates. d_e corresponds to the Hertzian contact deforma-

tion. $R_{x,y}$ corresponds to the contact radii of the rolling element and the bearing raceway. In the Etrul-Grubin analytical approach, $\hat{H}^*$ corresponds to the point where the dimensionless pressure and coordinate equal 0 and $\hat{H}$ corresponds to the dimensionless pressure using the Etrul-Grubin analysis. X is the dimensionless coordinate along the bearing raceway.

$$h(x,y) \approx h_o + \frac{x^2}{2R_x} + \frac{y^2}{2R_y} + d_e(x,y) \tag{4.1}$$

$$\hat{H} - \hat{H}^* = |X|(X^2-1)^2 - \ln[|X|+(X^2-1)^{(}1/2)] \tag{4.2}$$

Bearings are greased regularly and at predetermined amounts to maintain the minimum grease level inside the bearing. The general equation to calculate the amount of grease for a specific bearing is shown in Equation 4.3. L_r is the lubrication amount in grams, D is the outer diameter of the bearing, and B_w is the bearing width. The calculation for determining the intervals between bearings is given in Equation 4.4 from the Noria Corporation [482]. T represents the time in hours between lubrication. K is a product between all the correction factors shown in Table 4.9. n is the bearing speed in RPM, and d is the bearing bore diameter in mm. The bearing lubrication regime is determined for different manufacturing scenarios using Equation 4.3 and 4.4. Not following the regime or using a different lubrication than needed could lead to early bearing degradation. It can be seen from the calculations that deviations in the bearing operating environments can lead to changes in the lubrication pattern. If these are not accounted for and the timing is missed long enough, the bearing begins degradation earlier than the L_{10} life calculation.

$$L_r = DB_w * 0.005 \tag{4.3}$$

$$T = K * [(\frac{14000000}{n * d^{0.5}}) - 4 * d] \tag{4.4}$$

Even if the bearing is well-lubricated, contamination can enter through the bearing seals and disrupt the lubrication pattern. Contamination is described as any foreign substance not expected in the lubrication. Hence, the contamination can either be a solid or liquid. However, as these substances enter the raceway, they disrupt the lubrication film and increase the metal-on-metal contact, resulting in the skidding and sliding that is caused during the bearing wear. One particular

Table 4.9: Bearing Lubrication Interval parameters

Condition	Average Operating Range	Correction Factor
Temperature (Ft)	Housing below 150°F	1.0
	150°F to 175°F	0.5
	175°F to 200°F	0.2
	Above 200°F	0.1
Contamination (Fc)	Light, non-abrasive dust	1.0
	Heavy non-abrasive dust	0.7
	Light abrasive dust	0.4
	Heavy, abrasive dust	0.2
Moisture (Fm)	Humidity below 80%	1.0
	Humidity between 80% to 90 %	0.7
	Occasional condensation	0.4
	Occasional water on housing	0.1
Vibration (Fv)	Less than 0.2 ips	1.0
	0.2 to 0.4 ips	0.6
	above 0.4 ips	0.3
Position (Fp)	Horizontal bore centerline	1.0
	45°bore centerline	0.5
	vertical centerline	0.3
Bearing Design (Fd)	Ball Bearings	10
	Cylindrical and needle roller bearings	5.0
	Tapered and spherical roller bearings	1.0

problem is the overrolling of the particles that eventually cause dents in the bearing surface. Dwyer-Joyce [480] tested several different contamination particles to estimate the indentation size. These indentations accumulate as the amount of contamination enters into the bearing. Poddar *et al.* [87] and Maru *et al.* [483] tested the effect of contamination in bearing lubrication finding that the vibration level increases as the amount of contamination increased. The vibration increase also correlated with the increased scratches and indentations on the bearing surface. Hariharan *et al.* [484] demonstrated similar phenomena in grease lubrication; however, it is noted that the dispersion of contaminants is not necessarily equal due to the variation in viscosity between the two materials.

Grease contamination is captured in the bearing life calculation using the ISO reliability factor coefficient (α_{ISO}) from ISO 281 [334]. The equation for calculating one of the ISO reliability value curves is shown in Equation 4.5. κ is the reference viscosity ratio based on the viscosity related to the bearing dimensions and the bearing actual viscosity. e_c is the bearing contamination factor dependent on the number of particles, the reference viscosity ratio and the bearing pitch diameter. C_u is the bearing fatigue load limit, and P is the bearing dynamic load limit. The reliability factor is then multiplied against the normal L_{10} life equation to determine the expected bearing life under

contaminated conditions. Hence, as the bearing contamination increases, the related bearing life decreases.

$$[htb!]\alpha_{ISO} = 0.1[1 - (2.5671 - \frac{2.2649}{\kappa^{0.054}})^{0.83}(\frac{e_C C_u}{P})^{1/3}]^{-9.3} \tag{4.5}$$

It was determined to split the tests for lubrication and contamination damage into two categories. Lubrication would test the bearing performance under different lubrication amounts in the bearing. The bearing contamination cases would test the bearings under different contamination scenarios. The amount of lubrication removed from the bearing was based on Equation 4.3. The contamination amounts were initially based on the expected ISO contamination levels for oil from ISO 4406 [485]. The amounts are further explained in Section 4.2.3, which details the procedures for inducing bearing damage.

4.2.2.3 Bearing Assembly Defect: Brinelling

Dents and nicks form in the bearing raceways from assembly and operating errors. According to the SKF documentation [7], the most common problem occurs when forcing the bearing on an incorrectly fitted shaft. It is recommended to cool and heat the bearing to fit the shaft; however, the incorrect temperature does not cause the metal to contract enough to allow the bearing to fit onto the shift. Another common incorrect installation practice is unevenly pressing the bearing onto the shaft. An uneven press could cause nicks and dents to form on the raceway due to incorrect pressure. Another possible implementation is dropping the bearing during production or installation. The drop can send a shock through the bearing, leaving an indentation on the bearing surface.

It is possible to induce defects using hardness testers to resemble the expected defect size [486]. One of the common methods is using hardness testing equipment to induce the defect and measure the defect formation as time progresses. Ueda *et al.* [487] used both hardness testing equipment and contamination to induce a defect in a bearing test rig to induce the equipment flaking behavior. However, these defects are more meant to represent the plastic deformation case, which represents severe mounting errors and, more so, the shock loads affecting the equipment.

The method used is a drop test of the bearing on a metallic surface from a predetermined height. The action is meant to mime a situation where a maintenance engineer installs the bearing and may inadvertently drop the bearing. The shock to the bearing may cause a defect on the raceway

due to the shock of the rolling elements on the bearing inner race and outer race. Figure 4.13 shows the fixture for the bearing assembly defects. This case is termed the "Brinelling case" due to the similarities to the hardness test of the same name, but it is not an actual Brinell test. The bearing drops represent the potential for improper bearing handling. The different heights symbolize the potential heights that the bearing could fall. The drops could represent a small shock load to impact the bearing.

Figure 4.13: Fixture for ensuring repeatable drop testing

4.2.2.4 Damage Implementation: Summary

Extracted Information: From this step, four different methods are identified to apply damage to the bearing. For the fatigue case, defects are induced by engraving minor point defects using a tabletop CNC mill. Rather than using a drill, the flaws are generated using an abrasive engraver to create a rougher damage edge than the sharp and clean grooves. The lubrication defect implies starving the bearing from the bearing and increasing the metal-on-metal contact. The contamination

damage uses abrasive particles to induce defects along the bearing raceway, simulating the defect damage on the bearing surface. For the assembly and mounting defects, the bearing case investigates drop damage to the bearing raceway to simulate an installation error.

4.2.3 Determine the Damage Propagation: Bearings

The damage progression step was determined by each failure based on considerations from prior literature. The tables here represent the empirical progression for each of their failure modes. Other failure progressions were tested in the course of the research work and are labeled appropriately with the appropriate figures and analysis.

For the fatigue case, the damage was measured based on the defect size in terms of length along the bearing raceway and the bearing tool. A deburring tool that was 2.38 mm (3/32 in) in diameter was used to inflict damage on the bearing. Due to the spherical nature of the tool, the damage area was approximated based on the tool. The depth was approximated based on the force applied to the tool edge. Table 4.10 contains the different damage stages inflicted on the bearing case. The damage stages listed "Stage 1", "Stage 2", and "Stage 3" were tested on all bearings. The damage stages labeled "Stage 1 6205" and "Stage 2 6205" were conducted with a hand engraving tool to see if there was a difference between a point or line defect along the raceway. The bearing CNC length did not control the depth; hence, the damage percentage was conducted based on the percent surface area damaged.

Table 4.10: Bearing Fatigue Defect Cases

Damage Stages	Defect Length	Depth
Stage 1	1 mm	0.1 mm
Stage 2	3 - 5 mm	0.1 mm
Stage 3	5 - 10 mm	0.1 mm
Damage Stages	Damage Style	Percentage damage
Stage 1 6205	Point defect	0.2%
Stage 2 6205	Line defect	2.0%

The measures are based on the calculated grease amount for the bearings under expected lubrication cases. When opening the bearings to apply the grease, it was noticed by weighing the bearings that the bearings appeared underfilled about the expected grease amount. Table 4.11 contains the lubrication cases as a percentage of the original grease amount. About the original amount of grease in the bearing, though, Stages 1 and 2 for each case represent over-greased scenarios

in every case. Stage 3 for the 6205 and 6206 was an under-greased scenario, whereas Stage 3 for the 6207 bearing case represented a case for the proper amount of grease. The likelihood of bearing running without grease in an industrial application is low; hence, it was not included as part of the final test plan. It was considered, however.

Table 4.11: Bearing Lubrication Defect Cases

Damage Stages	Percentage	Lubrication Amount
6205 Stage 1	75%	2.92 g
6205 Stage 2	50%	1.8 g
6205 Stage 3	25%	0.97 g
6206 Stage 1	75%	3.72 g
6206 Stage 2	50%	2.48 g
6206 Stage 3	25%	1.24 g
6207 Stage 1	75%	4.59 g
6207 Stage 2	50%	3.06 g
6207 Stage 3	25%	1.53 g

The contamination case is based on the ISO oil codes for contaminated cases. However, due to the grease viscosity, the particle counts are sometimes skewed and unevenly distributed during the analysis. Hence, these values only represent the starting amount of contamination relative to the bearing size. Table 4.12 contains the number of grease particles per gram and the total amount of contaminated particles in grams. The amount of particles is calculated based on the particle density relative to the particle size. A spherical shape was assumed; particle amount changes based on the volume assumed. The diameter, though, was assumed by the grit size of the particles at 53 μm (F230 on the FEPA (Federation of European Producers of Abrasives) scale). The particles were alumina oxide as used by Dwyer-Joyce in their work [480]. Dwyer-Joyce used the contamination grit size to reference the particle size. Poddar *et al.* [87] used a similar range of sizes. Note that while ISO codes are used to identify the particle amounts relative to the grease amount, generally, these are employed with smaller grit sizes in mind. That being said, it is not unusual to notice grit sizes of the tested size, as Dwyer-Joyce noted. Additional contamination stages were conducted for earlier 6205 bearing tests.

The bearing drop tests are conducted based on the number of drops at certain heights. Table 4.13 contains the different damage stages. The OG stages are only associated with an earlier 6205 case. The increasing damage height can correspond to different impacts the bearing may undergo. However, since no force is applied, the impact of the bearing on the ground is considered due to gravity each time. However, the energy with each strike changes proportional to the change in

124

Table 4.12: Bearing Contamination Defect Cases

Damage Stages	Particles per gram	Contamination (g)
6205 Stage 1	10,000	0.012 g
6205 Stage 2	100,000	0.125 g
6205 Stage 3	250,000	0.250 g
6206 Stage 1	10,000	0.0159 g
6206 Stage 2	100,000	0.159 g
6206 Stage 3	250,000	0.318 g
6207 Stage 1	10,000	0.019 g
6207 Stage 2	250,000	0.192 g
6207 Stage 3	500,000	0.384 g

height.

Table 4.13: Bearing Drop Testing

Damage Stages	Damage Height	Number of Drops
Stage 1 (OG)	0.5m	x2
Stage 2 (OG)	1.0m	x2
Stage 3 (OG)	1.5m	x2
Stage 4 (OG)	2.0m	x2
Stage 1	0.5m	x5
Stage 2	1.0m	x5
Stage 3	1.5m	x5
Stage 4	2.0m	x5

Extracted Information: The different damage progression methods are presented for each case with the appropriate methods of damage to denote the change in equipment health over time. The different damage stages provide the damage label for any bearing data labeled.

4.2.4   Select the Data Acquisition System: Bearings

The bearing data acquisition system for testing considers three data types: vibration, acoustic, and temperature. These data types are commonly used in bearing condition monitoring strategies. Two vibration sensors are placed in the x- and y- directions. The vibration sensors are VSA001 from IFM[10]. The sampled frequency from the sensors was at 50kS/s. The sampling frequency exceeded the Nyquist frequency for the sensor harmonic levels by 8x. The Nyquist frequency was configured based on the maximum bearing frequency expected from the test stand. An infrared temperature sensor[11] was placed in the axial direction (z-direction). The sensor measured the ambient and infrared temperature around the bearing fixture at an approximate sampling rate of 0.1

---

[10]Vibration Sensor Information: https://www.ifm.com/us/en/product/VSA001
[11]Temperature Sensor Information: MLX90614ESF-DCA-000-SP at mouser.com

kS/s. For bearing failures, the temperature change can happen rapidly rather than sampling periodically; hence, the rapid sampling rate for the temperature sensors. Figure 4.14 shows the vibration and temperature sensors concerning the bearing test blocks. The solid blue arrow references the axial direction, and the dashed blue arrow references the radial horizontal direction. The solid orange circle corresponds to the vertical radial vibration sensor, and the dashed orange circle denotes the horizontal radial sensor. The green sensor denotes the temperature sensor pointed at the bearing to measure the temperature.
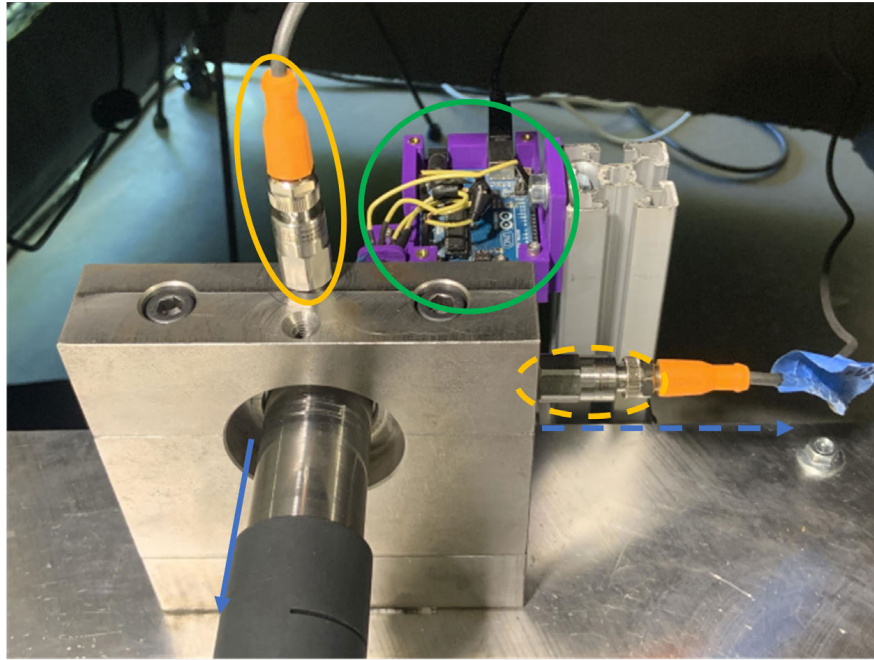
Figure 4.14: Vibration and temperature sensor layout for bearing testing

A microphone was placed in the horizontal radial direction to corroborate the frequencies gathered from the vibration testing [12]. The data measured from the analysis is scaled concerning the db and gain of the microphone. The sampling frequency was 44.1 kS/s, which exceeded the test bearings' Nyquist frequency. The microphone gain and scale are meant to remain constant during testing. Figure 4.15 shows the microphone layout concerning the bearing test stand. The end of the microphone was placed approximately 10 inches - 1 foot from the test bearing. The distance configured between the microphone and the test bearing was constrained by the overall safety cage around the bearing test stand. The microphone may still pick up noise around the test stand;

---

[12]Microphone Information: https://rode.com/en/microphones/shotgun/ntg3

insulation was placed around the test stand to reduce the amount of potential noise. The variation in noise from surrounding activities causes variability in the test stand microphone data.

Figure 4.15: Microphone concerning the test bearing

Extracted Information: The bearing data acquisition system was configured based on the applications under test. The sampling rates were configured based on the maximum expected bearing frequency. Based on the documentation, the measurement ranges for the sensors were expected to cover the maximum possibility from the test stands.

4.2.5   Design the Surrogate System and Generate Data: Bearings

A contributing goal for this research is to determine how to scale laboratory data to a real-world bearing scenario utilizing informed learning. Prototyping is a common strategy in design; hence, a similar strategy is adopted in the formulation of the surrogate system design. The first consideration is to identify the primary elements of the real-world system. The second consideration is to consider and determine the physical quantities in the prototype design. The data generation occurs after the surrogate system design, where the labels are determined based on the damage progression and operating conditions.

For bearings, the primary physical considerations for each application consider the bearing geometry, the bearing speed, and the bearing load. The bearing geometry and speed determine the vibration frequencies based on the bearing defect equations. The bearing load and speed influence the vibration amplitude. Other physical quantities should also involve other rotational elements

127

on the shaft and surrounding equipment. Hence, a bearing test system should consider modular elements to allow for other rotational elements to be added to the surrogate system as needed.

Figure 4.16 shows the bearing test stand constructed for the data generation. The test stand is split into three levels: the testing level (top level), the load level (middle level), and the motor level (bottom level). The testing level comprises the test block, which holds the test bearings, and the support blocks, which hold the driving shaft. The loading level contains the loading arm that applies a load to the bottom of the bearing by adding weights to the back of the arm. The loading mechanism comprises a force sensor to verify the force applied to the bearing. The bearing test block is also placed on a separate rotating block to induce operational play during testing. The motor level contains the motor, which connects to the test level using the pulleys and v-belt. The motor is connected to a variable frequency drive to ensure variable speed during testing. The test stand during operation is placed inside a protective cage.

Four different bearings were considered for the analysis and data generation: 6204, 6205, 6206, and 6207 cases. The test blocks were precision machined to serve as housings for the bearings as a split housing to close around the bearing. Figure 4.17 shows the bearing test block configuration with the top removed for the split housing. The top block is what contains the vibration sensor during testing. The housings are configured to the outer diameters and represent a perfect fit for the bearing under test. The bearings are pressed onto test shafts that are machined to the respective inner diameter for each bearing. A mounting shoulder is applied for each shaft to ensure the press fit for each bearing. The test block in Figure 4.17 contains the 6204 and 6205 bearing sizes. A separate test block, designed similarly, is used for the 6206 and 6207 bearings. The bearing test shaft connects to a drive shaft connected to the motor using the pulley and v-belt system. The final configuration of the bearing test shaft includes a cap at the end of testing to ensure the bearing does not move off the shaft during testing.

The bearing data generation is broken down into tests and collections. Each test conducted consists of collections that occur at different speeds and loads. There are three set speed configurations (640, 1103, and 1687 RPM) and three set loads (520N, 814N, 1049N) configurations. Due to the variability in the tests, the speed and loads did not fluctuate by more than 2% of the original configuration. Three tests are performed: Steady State, Speed Varying, and Load Varying. The Steady State considers collections when the operating conditions remain the same throughout the collection. The Speed Varying collections consider changing speed conditions during the data ac-
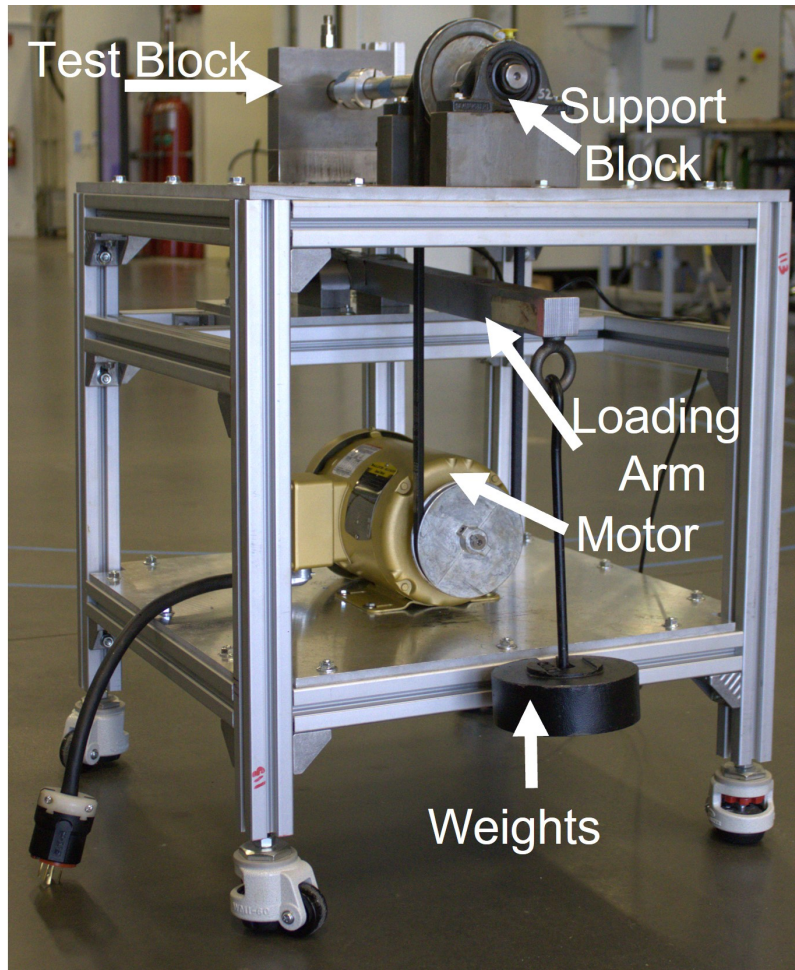
Figure 4.16: Bearing Test Stand

quisition procedure. The Load Varying collections consider instances where the load changes in the data. Figure 4.18 shows the generic configuration for a test at a set speed and load, incorporating the Steady State, Speed Varying, and Load Varying collections into the procedure. The warm-up periods are run between 5 to 10 minutes and the time changed based on the lack of variation between collections. The speed-varying collections considered changes in bearing speed up to 10 Hz (1 Hz corresponds to 30 RPM from the VFD). The load-varying collections considered changes of up to 90 - 100 N during the collections. The collection order was repeated for each speed and load configuration. Depending on the bearing, this procedure was repeated between 3 to 9 times for each test.

Extracted Information: The bearing test stand is constructed to control the bearing geometry,
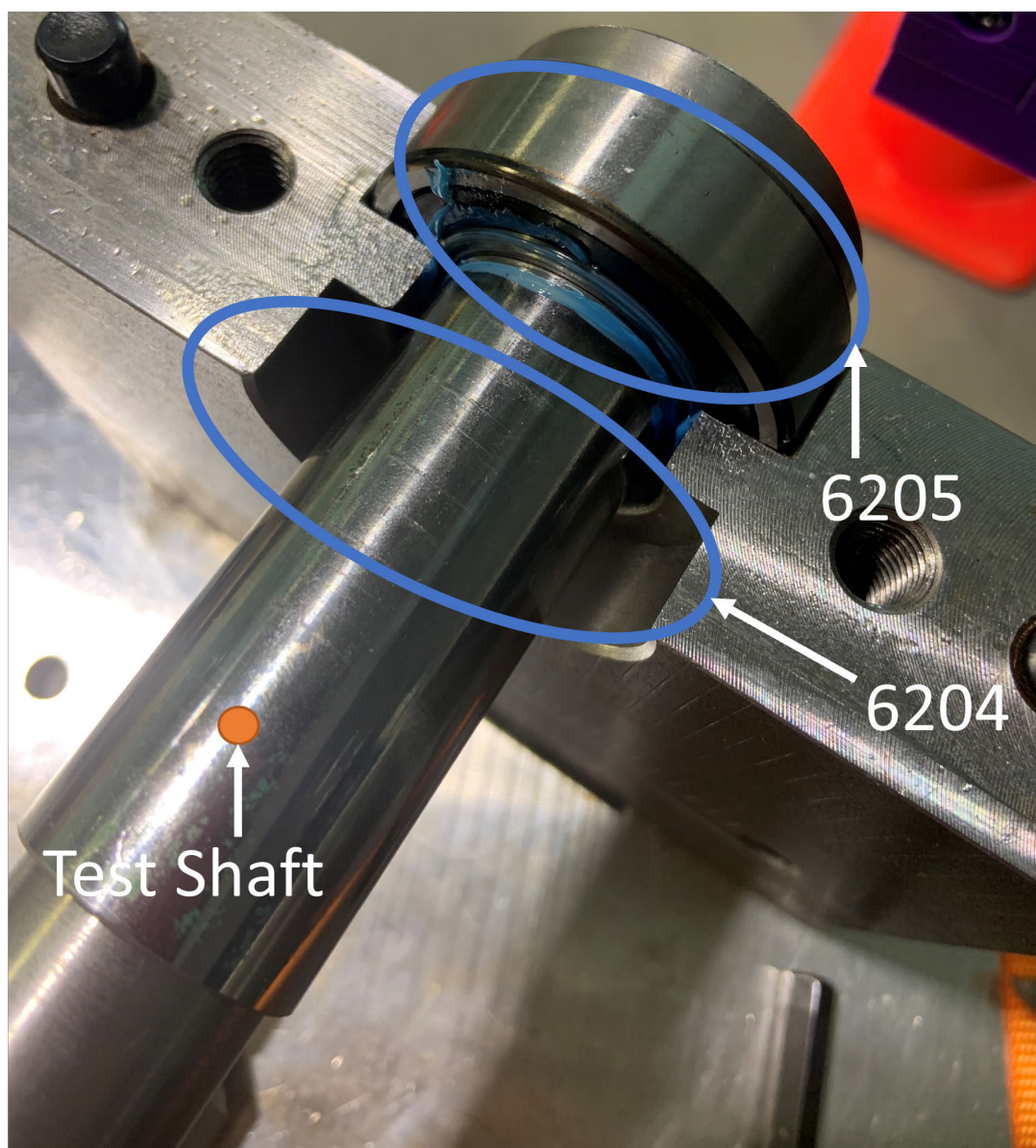
Figure 4.17: Test block configuration with the bottom block

speed, and load to provide relatability to the real-world system. The bearing test stands considered three test types: steady state, speed varying, and load varying. It was built in a modular format to ensure that other elements could be added depending on the real-world equipment.
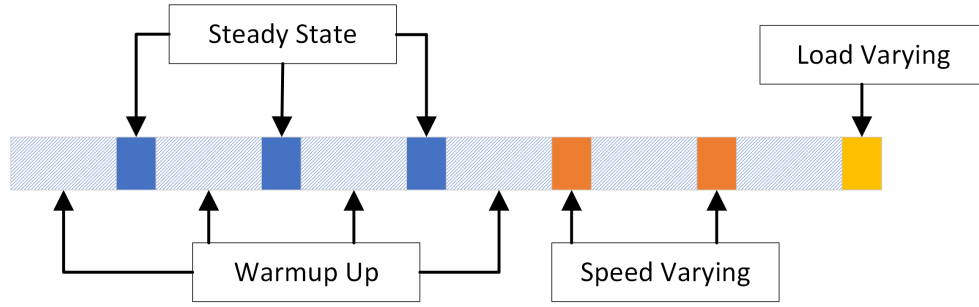
Figure 4.18: Example of bearing test procedure and types of tests

4.2.6   Verify and Transfer the Data: Bearings

For this step, the methods are introduced for verifying the generated data and transferring the bearing data between different applications. The analysis representations are further described in later chapters and sections. Section 4.3 covers verifying and analyzing the bearing data generated. Chapter 5 provides a more in-depth consideration of the methods (data-driven and physics-informed) used to transfer bearing data from differing applications. The concepts are briefly introduced in this section, though.

Three verification methods verify the bearing data generated from the test stand. The first method considers statistical features and groupings to determine if the bearing data is the same between similar classes and separable between different classes. ANOVA tests are conducted to determine if there is any statistical significance in any data overlap. These methods consider only the time series data, which are grouped based on the sampling frequency. Data windowing is considered in Chapter 6 to determine if there is a significant influence on the bearing data.

The second method considers the frequency measures for devolving the bearing data using Fourier transforms into frequency representations. Particular attention is assessed to the bearing defect frequencies shown in Equation 2.33 through 2.36. These frequencies are represented in both Hertz and Undulations Per Revolution (UPR). UPR represents the bearing frequencies without their corresponding speed component, allowing for a direct representation concerning load. As bearing damage increases, these frequencies should become more apparent in the bearing frequency analysis.

The third method utilizes Empirical Mode Decomposition (EMD) to devolve the bearing signal into the intrinsic mode functions (IMFs). EMD is an integral portion of the Hilbert-Huang Transform (HHT), which identifies the instantaneous frequencies of a sample utilizing the IMFs [488]. Huang *et al.* [488] theorized that EMD could determine the IMFs by using a sifting process

131

to ensure that a finite number of IMFs are returned to determine the instantaneous frequencies. An intrinsic mode function comprises two criteria:

i.) In the whole data set, the number of extrema and zero-crossings must either be equal or differ at most by one.

ii.) At any point, the mean value of the envelope defined by the local maxima and local minima is zero.

The IMFs represent the original signal and are defined as a near orthogonal representation. Generally, the first IMF is rejected as it contains primarily high-frequency noise. Zhao *et al.* [489] and Wu *et al.* [490] each used empirical mode decomposition in their rolling bearing analysis. Zhao *et al.* used EMD to devolve the bearing signal and then reconstruct the signal in the frequency domain while removing nonsensitive IMFs to bearing damage. Wu *et al.* used the HHT to identify the change in spectral energy as the bearing degrades over time. For this analysis, the generated bearing data IMFs are compared to see how they compare to other collected bearings.

For the transfer approaches back to the real system, two new approaches are formulated to transfer data from the surrogate system to the real-world application. The first method is termed the Spectral Augmentation (SA) method. The method combines data augmentation with the bearing defect frequencies to transfer the failure components to baseline data. Theoretically, the baseline data with the failure components will closely match the bearing failure signals. The second method uses bearing physical modeling to determine vibration limits to remove potential noise for transferring bearing data. Scaling factors are determined based on the change in physics modeling representation to move bearing data between different applications. The physics analysis is tested with the baseline data, and observational limits are theorized with the fault data and simulation of rolling element point defects. Chapter 5 covers these methods in further detail later.

Extracted Information: Three different data analysis techniques are conducted to verify the bearing data. A physics-informed method and physics-based methods are introduced for data transfer in Chapter 5.

4.2.7 Section Recap: Failure Methodology Design

This section determines the method for generating bearing damage, the design and creation of the surrogate system for bearings, and the selection of the analysis techniques for bearings. At this

point, the data generation provides a wide, diverse data pool for further exploration of the bearing failure modes under different conditions. At this point, the third research question introduced at the beginning of the section cannot be answered. However, this section provides the means to answer the question in the following section.

4.3   Data Generation: Bearings

The generated bearing data using the methodologies from Section 4.2 are compared against each other to investigate RQ1.3:

i.) Research Question 1.3: How much data are needed to train a bearing condition monitoring system for diverse failure modes?

In this section, the techniques briefly introduced in Section 4.2.6 describe the changes in the bearing data under the different class conditions. Table 4.14 contains the other tests conducted for each load and speed condition on the bearing system. The "s" and "k" refer to the different manufacturers of the bearings. Note that this did not appear to skew the results in the analysis and is just a notation. The "base" refers to whether the bearing test conducted was a baseline test, "Brinell" denotes the bearing was used for a Brinell test, "lube" corresponded to the lubrication tests, and "fatigue" was for the engraving tests. The different types of collections used in each test are denoted as follows: ss means "steady state", sv means "speed varying", and lv means "load varying".

Table 4.14: Bearing tests conducted for each set of operating conditions and bearing

Speed → Load ↓	640 RPM	1103 RPM	1697 RPM
520N	2k6205 (base, ss, sv), 2k6205 (lube, ss, sv), 3k6205 (base, ss, sv), 3k6205 (brinell, ss, sv), 4k6205 (base, ss, sv), 5k6205 (base, ss, sv), 6k6205 (base, ss, sv), 3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 2k6205 (base, ss, sv), 2k6206 (brinell, ss, sv), 3k6206 (base, ss, sv), 3k6206 (lube, ss, sv, lv), 4k6206 (base, ss, sv, lv), 5k6206 (base, ss, sv, lv), 5k6206 (cont, ss, sv, lv), 2k6207 (base, ss, sv), 3k6207 (base, ss, sv), 3k6207 (cont, ss, sv, lv), 4k6207 (base, ss, sv), 4k6207 (brinell, ss, sv), 5k6207 (base, ss, sv), 6k6207 (base, ss, sv)	2k6205 (base, ss, sv), 2k6205 (lube, ss, sv), 3k6205 (base, ss, sv), 3k6205 (brinell, ss, sv), 4k6205 (base, ss, sv), 5k6205 (base, ss, sv), 6k6205 (base, ss, sv), 3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 12k6205 (fatigue, ss, sv, lv) 2k6205 (base, ss, sv), 2k6206 (brinell, ss, sv), 3k6206 (base, ss, sv), 3k6206 (lube, ss, sv, lv), 4k6206 (base, ss, sv, lv), 5k6206 (base, ss, sv, lv), 5k6206 (cont, ss, sv, lv), 12k6206 (fatigue, ss, sv, lv), 2k6207 (base, ss, sv), 3k6207 (base, ss, sv), 3k6207 (cont, ss, sv, lv), 4k6207 (base, ss, sv), 4k6207 (brinell, ss, sv), 5k6207 (base, ss, sv), 6k6207 (base, ss, sv), 12k6207 (fatigue, ss, sv, lv)	2k6205 (base, ss, sv), 3k6205 (base, ss, sv), 3k6205 (brinell, ss, sv), 4k6205 (base, ss), 5k6205 (base, ss), 6k6205 (base, ss), 3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 12k6205 (fatigue, ss, sv, lv) 2k6205 (base, ss, sv), 2k6206 (brinell, ss, sv), 3k6206 (base, ss, sv), 3k6206 (lube, ss, sv, lv), 4k6206 (base, ss, sv, lv), 5k6206 (base, ss, sv, lv), 5k6206 (cont, ss, sv, lv), 12k6206 (fatigue, ss, sv, lv) 2k6207 (base, ss, sv), 3k6207 (base, ss, sv), 3k6207 (cont, ss, sv, lv), 4k6207 (base, ss, sv), 4k6207 (brinell, ss, sv), 5k6207 (base, ss, sv), 6k6207 (base, ss, sv) 12k6207 (fatigue, ss, sv, lv)
814N	3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 11s6205 (base, ss), 7s6205 (fatigue, ss), 12s6206 (base, ss, sv), 13s6206 (base, ss, sv)	3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 11s6205 (base, ss), 12s6206 (base, ss, sv), 13s6206 (base, ss, sv)	3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 11s6205 (base, ss), 12s6206 (base, ss, sv), 13s6206 (base, ss, sv)
1049N	3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 8s6205 (cont, ss), 11s6205 (base, ss, sv), 14s6205 (base, ss), 12s6206 (base, ss, sv), 13s6206 (base, ss, sv)	3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 8s6205 (cont, ss), 11s6205 (base, ss), 14s6205 (base, ss), 12s6206 (base, ss, sv), 13s6206 (base, ss, sv)	3s6205 (cont, ss), 4s6205 (base, ss), 4s6205 (brinell, ss), 5s6205 (base, ss), 7s6205 (base, ss), 7s6205 (fatigue, ss), 8s6205 (base, ss), 8s6205 (cont, ss), 11s6205 (base, ss), 14s6205 (base, ss), 12s6206 (base, ss, sv), 13s6206 (base, ss, sv)

4.3.1 Baseline Operating Conditions Differences

The baseline bearing data were analyzed using Fast Fourier Transform (FFT), Time-based statistical measures, and Empirical Mode Decomposition. Fast Fourier Transform is a common frequency analysis measure created by Cooley *et al.* [491] to originally monitor missile launches. Using the FFT, it is possible to change from time-series data to frequency-based data rapidly and easily. Time-based statistical measures capture and detail changes to specific moments within the data, such as changes in the root-mean-square, variance, skewness, and kurtosis. Finally, the empirical mode decomposition method is used as a representation to provide an understanding of the measures that change with respect to speed.

For the initial baseline characterization, the bearings were run at for varying lengths for 2 and a half hours to three hours at varying speeds and loads. It should be noted that data were collected in two different locations: Clemson Vehicle Assembly Center and the Clemson Graduate Education Center. Interestingly, it appeared to affect the data collection in the low-frequency range below 500 Hz for the vertical accelerometer data. Several reasons could affect the vibration data collected between the two locations. Mechanical components could become worn out during testing; a belt could begin to loosen or lose grip over time; a difference in the surrounding equipment could affect the vibration readings; the environmental conditions are not conducive for vibration testing. All of these factors could cause a difference in the bearing data generation. Hence, these factors require careful documentation to reduce their impact on baseline data generation. It should be noted that the horizontal bearing data was not as significantly impacted by the change of location.

Due to the lower variation in the horizontal bearing vibration, it is more likely that there was a faulty component during the baseline data generation. Despite changing the components and making modifications to reduce the test stand noise, the level of noise was not diminished. The horizontal bearing data does not appear to change with respect to the noise. It could mean that the noise is created due to the loading mechanism acting on the bearing. The variation is further explained in the frequency data for potential possible variations.

Figure 4.19 shows the grouping of the baseline bearings based on the vibration variance along the x-axis and the vibration RMS along the y-axis. The comparison did not consider two baseline bearings (Bearing 4 and Bearing 7) due to errors in the test procedure. For each case, the grouping of data points appears with higher RMS values associated with higher variance values. In

the 640 and 1103 cases, the data appear as a tighter grouping in the run-to-run case. For the 1687 RPM case, there is an increased variation between each baseline bearing. It is possible that the increase in speed causes additional vibrations to come through the test stand. Despite the variation in the 1687 RPM case, it is possible to see that there are clusters that form between each set of operating conditions with some overlap between the different bearing characteristics. The averages for each case are shown in Table 4.15. The RMS and variance values were found to increase at different rates relative to the change in speed. The rapid increase in relation to the variance could stem from additional harmonics caused by the increase in the operation speed.

A silhouette test was conducted to assess the level of overlap between the different clusters created with the different operating conditions. The silhouette score considers the overlap between the two clusters. Traditionally, the score interprets how well the clusters are defined without overlap between the other clusters [492]. The silhouette score is assessed over a range of -1 to 1. If the score is closer to 1, then it is considered that the values form well-defined clusters with little overlap. If the silhouette is closer to -1, then labels are flipped for the data, but still remain well defined. If the score is closer to zero, then the clusters are not well defined and contain a large amount of overlap between samples. A further explanation of the silhouette score is provided in Appendix D.

For the RMS feature, the 640RPM and 1103RPM cases were found to have a silhouette score of 0.247. The silhouette score between the 1103RPM and 1687RPM case was measured at 0.542. For the variance feature, the silhouette score was at 0.44 between the 640 and 1103 RPM case. The 1103 and 1687 RPM case had a silhouette score of 0.14. The silhouette scores imply that the overlap between the different cases was either moderate or low between the different cases. When these features are considered in a group, the 640 and 1103 RPM case have 0.512 overlap between the different data cases. The 1103 and 1687 RPM case had a silhouette score of 0.16. Hence, there was a lower amount of overlapping features between the 1687 RPM case and the 1103 RPM case, then there was with the 640 RPM and 1103 RPM case.

Bearing Condition	Average RMS	Average Variance
Baseline, 640 RPM	0.220 $\frac{m}{s^2}$	0.052 $(\frac{m}{s^2})^2$
Baseline, 1103 RPM	0.524 $\frac{m}{s^2}$	0.318 $(\frac{m}{s^2})^2$
Baseline, 1687 RPM	0.917 $\frac{m}{s^2}$	0.882 $(\frac{m}{s^2})^2$

Table 4.15: Averages of the RMS and Variance for the 6205 case for each set of operating conditions.

RMS is a data metric to describe the quadratic mean of a sample set. The variance describes
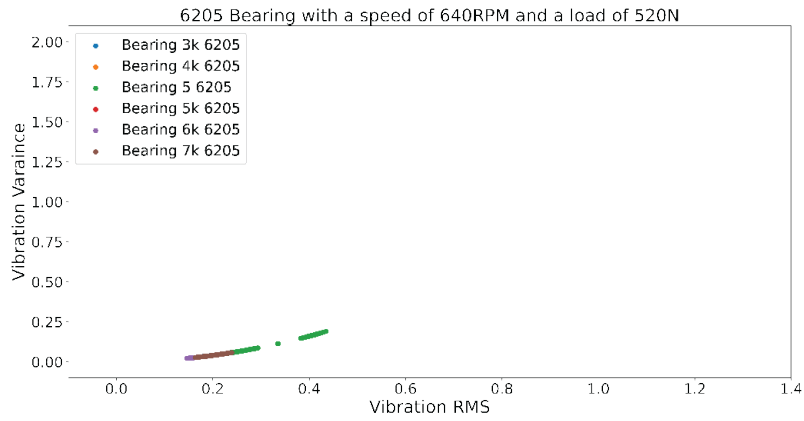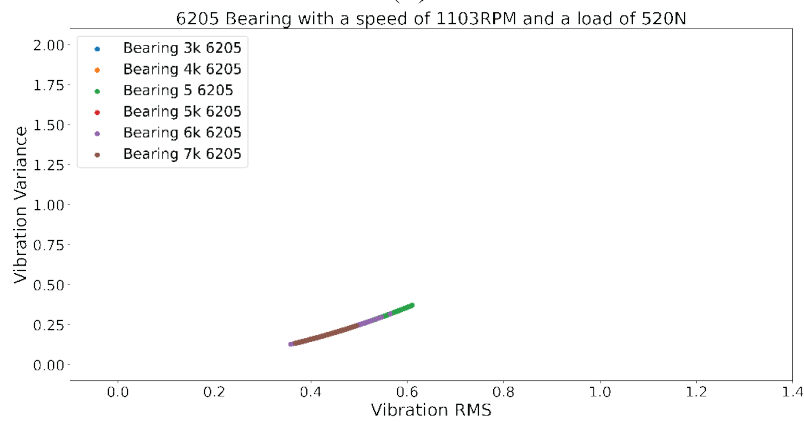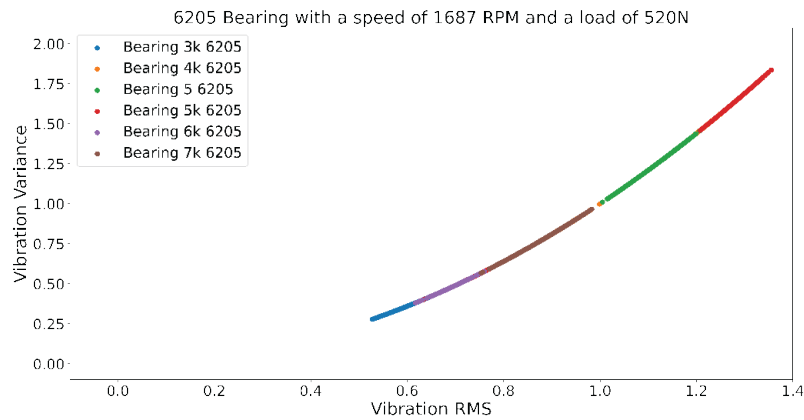
Figure 4.19: RMS and Variance for each baseline scenario in the 6205 case under changing speed conditions and constant load. (a) represents the bearing data colleced at 640 RPM, (b) represents the bearing data collected at 1103 RPM, and (c) represents the bearing data collected at 1687 RPM.

the spread of a number set in relation to the RMS. The Kurtosis considers the distribution with respect to the tail weighting relative to the sample set mean otherwise kown as the fourth moment of a distribution. The skewness of a dataset determines the data symmetry. From Mechefske [292], a baseline set of data is considered to have a near "normal" distribution. Depending on the application,though, the baseline distribution could change, such as those that are more transient systems. However, it is expected that the bearing data generated during the steady-state conditions mimic the normal distribution. Figure 4.20 shows, though, that there are some outlier elements that skew the data reporting. For future data analysis, it is possible to eliminate samples based on their normality or lack of normality. However, the limits for normality are subjective for using Kurtosis and skewness. For the range of skewness, the lowest acceptable limits found are between -0.5 to 0.5 for a normal distribution; however, -2 to +2 is also an acceptable range as found by George *et al.* [493] and Hair *et al.* [494]. In the ideal case, the For Kurtosis, George *et al.* considered the Kurtosis range between -2 and +2 for an accepted normality, whereas Hair *et al.* considered it at -7 to +7. Aminu *et al.* [495] indicated that Skewness and Kurtosis values above 3 and 10, respectively, could indicate a serious problem. For future analysis, the data points with Kurtosis values greater than 7 are removed from the dataset in the baseline case.

In addition to the steady-state condition monitoring case, transient systems were also considered in a load-varying case and speed-varying case. Figure 4.21 shows the load-varying case for the bearing baseline condition for the 6205 sized bearing. Only the 1687 RPM case is there a large variation between when the expected load is applied and an additional load is given. For the 1103 RPM case and the 640 RPM case, there is not any noticeable variation between the two different load cases. One potential cause for the lack of variation is the speed domination that could be occurring at lower frequencies. It is possible that applying a bandpass or highpass filter removes some of the speed domination in the bearing data.

Figure 4.22 shows the effects of the different bearing speeds and the variation during the course of the collection. The green circles mark instances where the speed is increasing. The black circles mark instances where the speed is decreasing. In this instance, the only outlier is the green data, which sees the steady-state variance shift with respect to the bearing during the test. In this instance, the bearing test conducted did not decrease back to the expected speed limit and was considered an anomaly due to the test procedure. For the frequency data, since the vibration data is tied heavily to the equipment operation, it is noted that the EMD method will isolate those
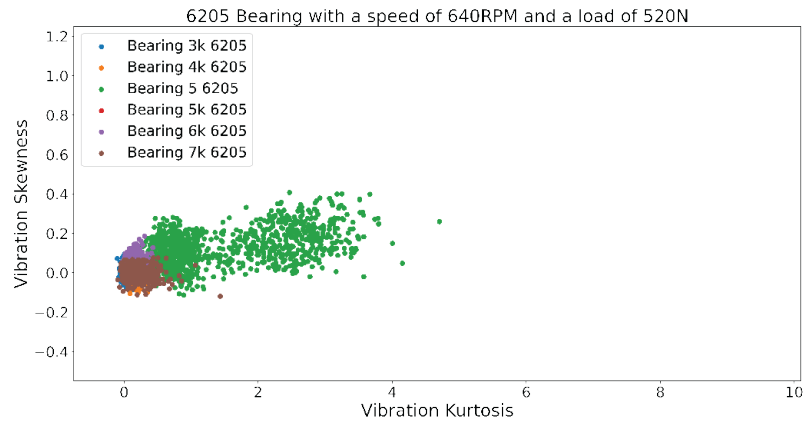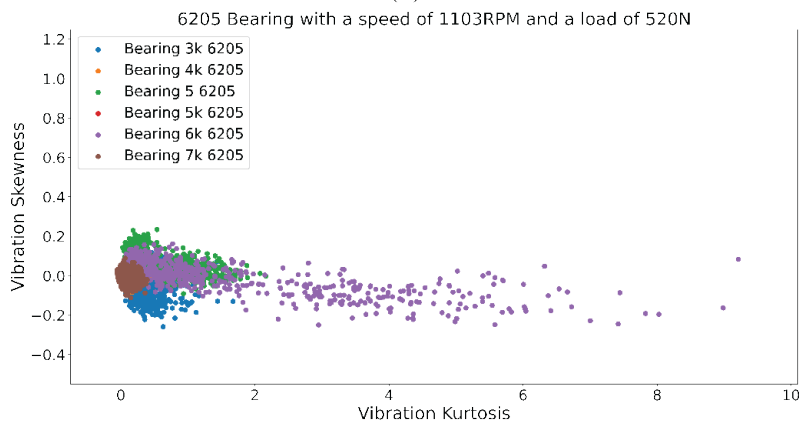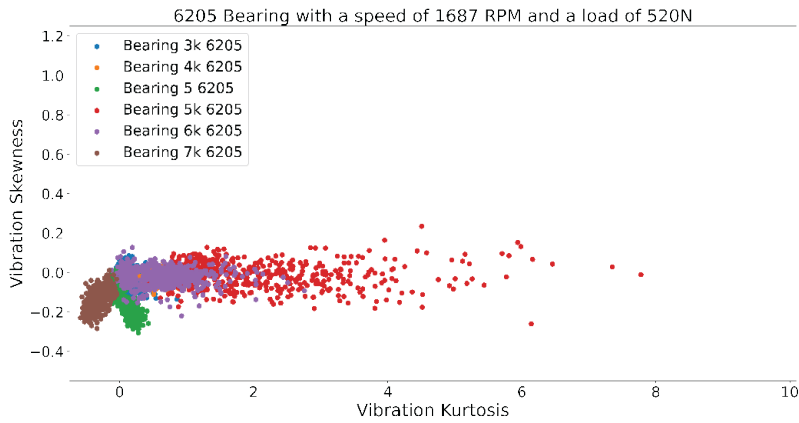
Figure 4.20: Kurtosis and Skewness for each baseline scenario in the 6205 case under changing speed conditions and constant load. (a) shows the statistical kurtosis and skewness for the 640 RPM data, (b) shows the statistical kurtosis and skewness for the 1103 RPM data, and (c) shows the statistical kurtosis and skewness for the 1687 RPM data.
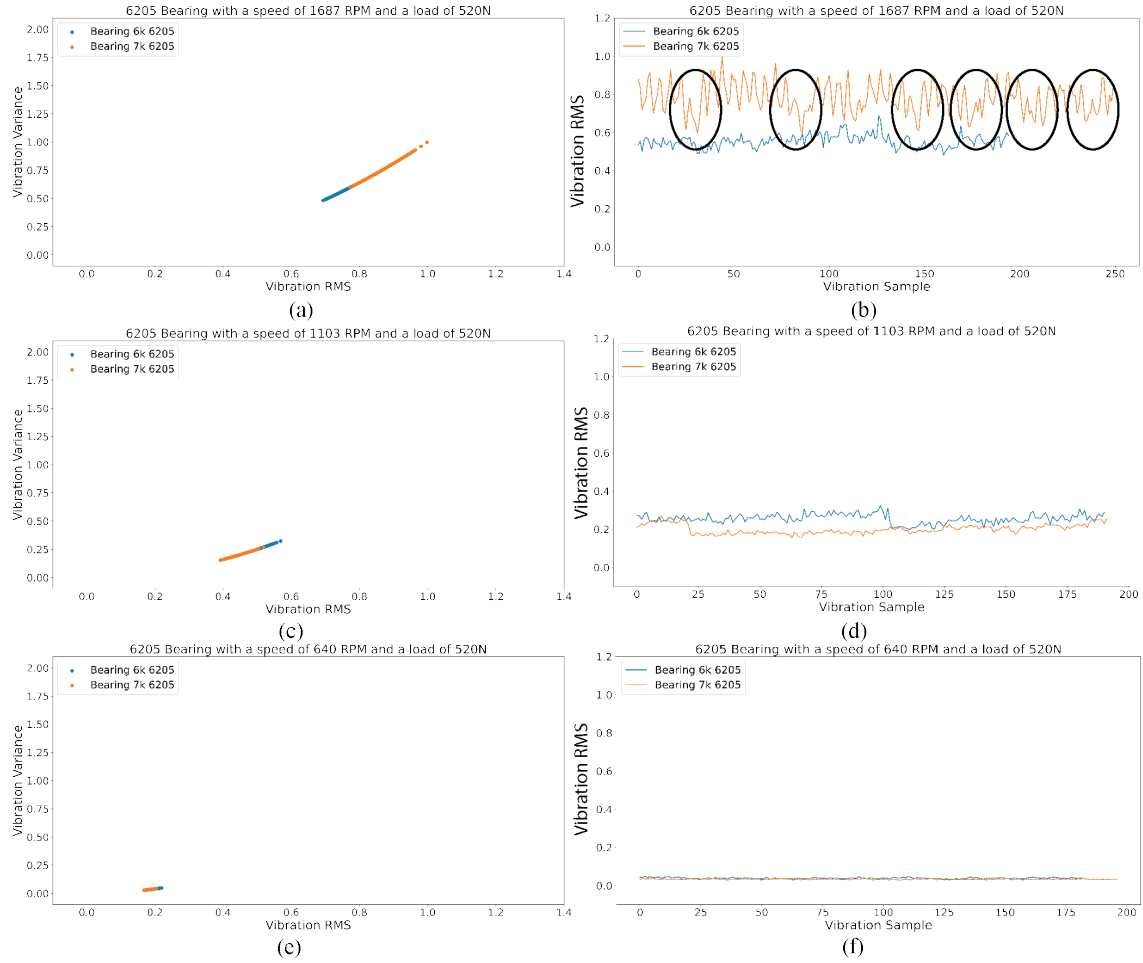
Figure 4.21: Load varying baseline raw vibration features for the RMS and variance cases for the 6205 data cases. (a-b) 1687 RPM case: (a) shows the RMS vs. Variance comparison, (b) shows the vibration sample set for the RMS feature. (c-d) 1103 RPM case: (c) shows the RMS vs. Variance comparison, (b) shows the vibration sample set for the RMS feature, (e-f) 640 RPM case: (e) shows the RMS vs. Variance comparison, (f) shows the vibration sample set for the RMS feature.

instantaneous frequencies attached at those speeds.

Figure 4.23 shows the variations when changing the testing locations in the vertical bearing data. A number of factors could have caused this inconsistency to occur and stay present in the bearing data. First, it is noted that most of this noise is low-frequency vibrations centered around the operating speed frequency. It is possible that an element of the test stand settled in and caused the misalignment as the testing progressed for over a year. However, changing the components other than the test block or mounted bearings did not solve the problem. A doubt for this observation also stems from the lack of variation in the horizontal vibration data as shown in Figure 4.19. An
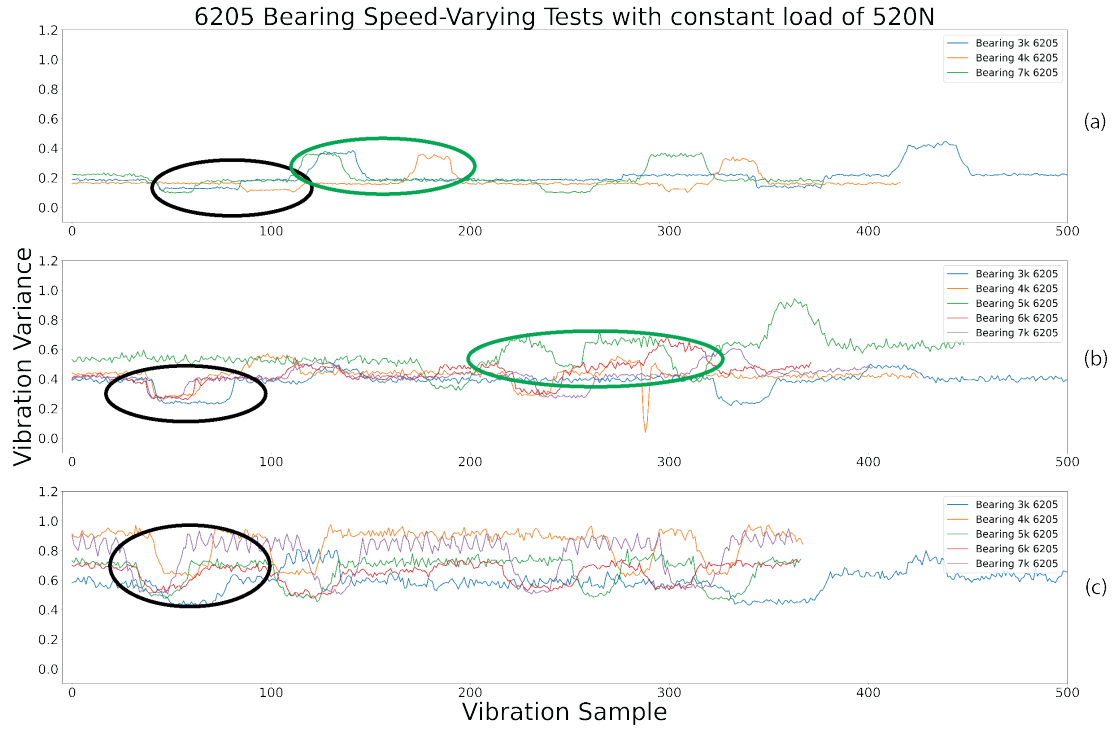
Figure 4.22: Speed varying test for the 6205 bearing under constant loads. (a) base speed of 640 RPM and then ± 10 Hz increments, (b) base speed of 1103 RPM and then ± 10 Hz increments, and (c) 1687 RPM base speed and ± 10 Hz increments

additional reason could be the loading arm mechanism impacting the bearing as it rotates in the block. Due to the rotation, the impacts would coincide with the bearing speed frequency. Another possibility could be environmental factors as a result of surrounding operating equipment. Vibration is susceptible to many different noise factors and has been found to be dependent on environmental factors, such as temperature and humidity [496]. Due to this, many failure mode comparisons occur on a bearing case-by-case comparison to the bearing baseline and a representative bearing baseline. However, the physics-informed methods discussed in Chapter 5 are investigated to see if removing these noise elements in the mechanical equipment is possible.

Figure 4.24 shows the horizontal vibration data for different bearing sizes. Each group has at least four bearings for each class for comparison. As the bearing size increases, the bearing vibration appears to increase relative to the bearing signature. For further reference, one 6207 bearings was discounted, because of errors in the test collection procedure for that bearing. There are no direct references for why the bearing vibration increases relative to the bearing size in the research literature. However, there are comparisons of equipment of different sizes that use bearings.

141
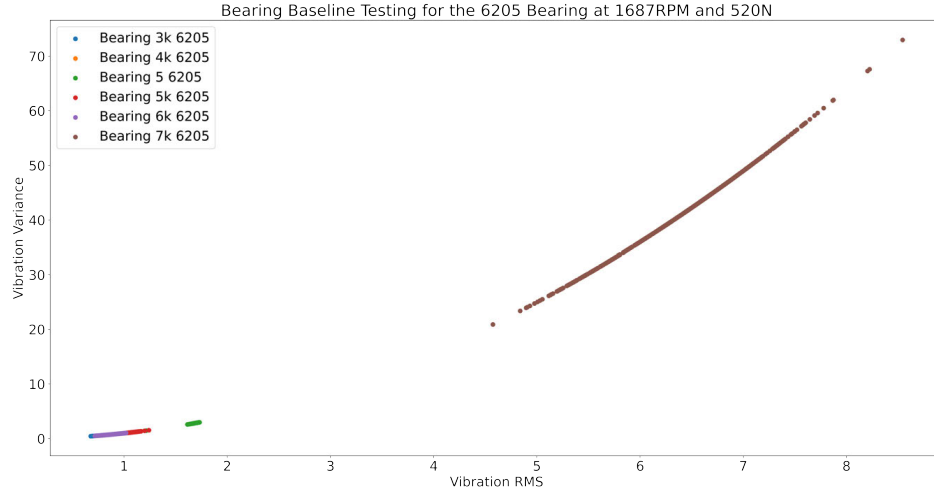
Figure 4.23: Example of inconsistencies of the vertical bearing features from RMS and variance. The variation exceeds the possible limits and is much greater than early data taken.

ISO 20816 is a series of standards that details the general vibration limits for non-rotating, rotating, and non-reciprocating equipment [389]. For example, ISO 20816-2 [497] provides the vibration limits for different equipment at certain speeds and horsepower. As the equipment speed or horsepower increases, the subsequent vibration readings increase. One could theorize this as corresponding to different bearing sizes that operate within larger equipment. Larger bearings generally handle larger loads at lower speeds, increasing the potential power that the bearing can handle in rotational equipment. Table 4.16 contains the corresponding parameters for the bearing speed and load ratings from two major bearing manufacturers, Koyo and SKF. These parameters are for deep groove ball bearings in their basic configuration and a standard clearance. As the bearing limiting speeds decrease, the bearing load ratings increase.

Table 4.16: Limiting speeds and dynamic load ratings for the 6205, 6206, and 6207 bearing sizes.

Bearing	Koyo Limiting Speed	SKF Limiting Speed	Koyo Dynamic Load Rating	SKF Dynamic Load Rating
6205	13000 RPM	18000 RPM	17.5 kN	14.8 kN
6206	11000 RPM	15000 RPM	24.3 kN	20.3 kN
6207	9200 RPM	13000 RPM	32.1 kN	27 kN

However, three different bearing sizes are tested under the same operating conditions. Another reason could stem from the difference in the bearing deformation as a result of a rolling element moving over a specific point. The formulation of bearing physics is further explained in Section 5.1.1.1; however, the probable reason for the difference in the vibration response should
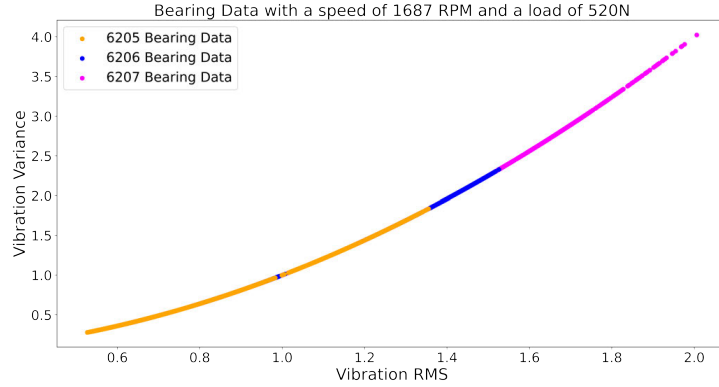
Figure 4.24: Example of variation between the 6205, 6206, and 6207 bearing data

stem from the deformation and rolling element stiffness acting on the bearing. As seen by Figure 4.25, the bearing rolling element stiffness increases as the bearing size increase. The dimensionless deformation quantity and the bearing curvature sum are the two parameters that correspond to the bearing rolling element stiffness. Figure 4.26 shows the subsequent effect on the bearing force due to the change in the bearing stiffness as a result of the bearing sizes. The subsequent increase in bearing force explains the difference in the vibration between the different bearing sizes. The bearing stiffness and force are based on equations from the following reference [498]. Different bearing equations are further examined in Section 5.1.1.1 to provide a comparison of the different estimated values in bearing parameters.
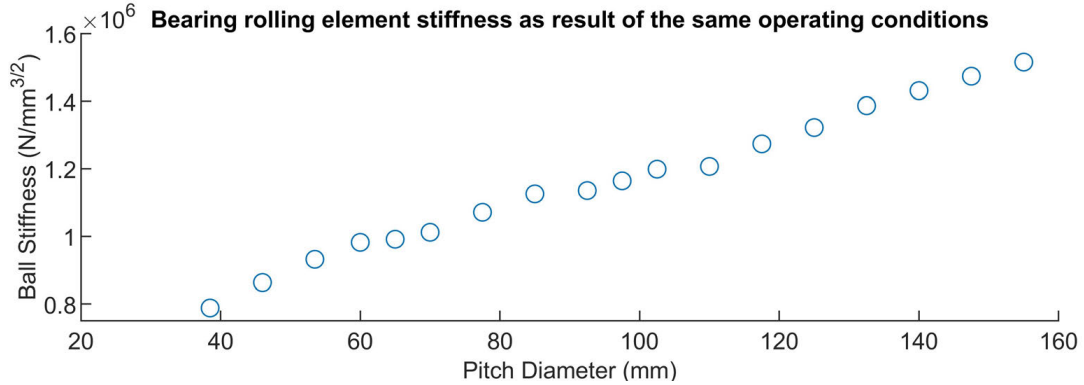
Figure 4.25: Bearing stiffness as a result of the bearing pitch diameter

Table 4.17 contains the average RMS and variance for the different bearing sizes. It is clear from Figure 4.24 that the bearing size does still overlap despite the difference in the average RMS and variance. Using the silhouette score, the 6205 and 6206 bearings had a score of 0.11 and a
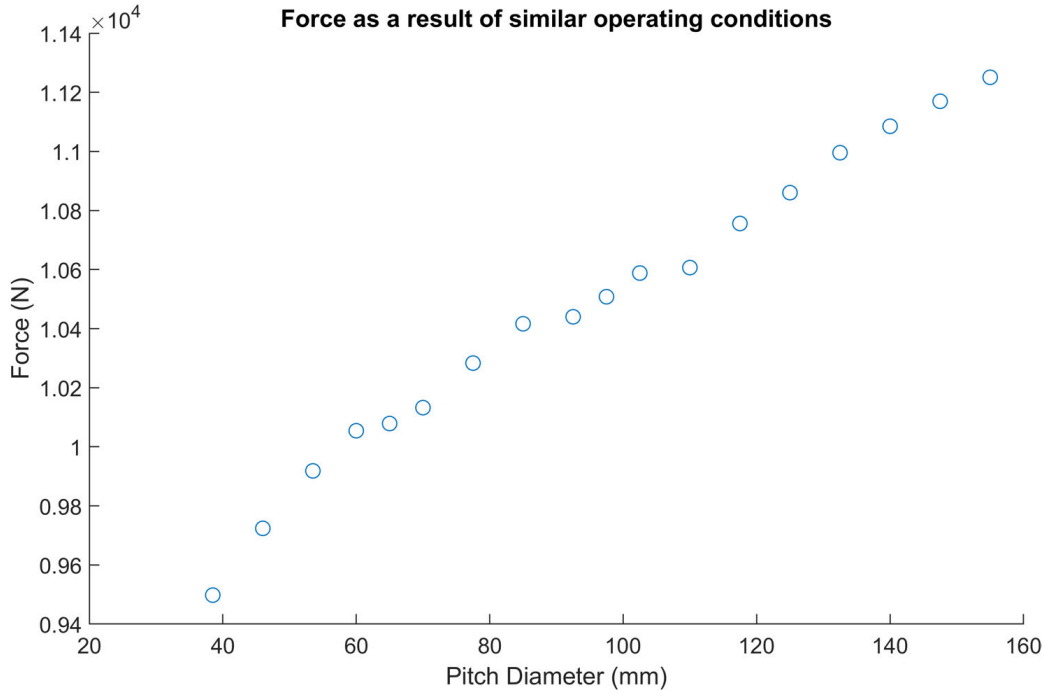
Figure 4.26: Bearing force striking as a result of the bearing stiffness and a predetermined defect size veruse the bearing pitch diameter

score of 0.03 between the 6206 and 6207 bearings. Due to the low silhouette score, these feature clusters overlap significantly, despite the larger differences in the variance score. Potential reasons for the difference in the bearing vibration and similarity could come from the geometric, kinematic, and dynamic differences between the different bearing sizes and testing. The similar and different kinematic, dynamic, and geometric differences between the bearings are further explained in Chapter 5.

Table 4.17: The average RMS and Variance for each bearing feature based on the bearing size

Bearing Size	Average RMS	Average Variance
6205	$0.917 \frac{m}{s^2}$	$0.882 \left(\frac{m}{s^2}\right)^2$
6206	$1.032 \frac{m}{s^2}$	$1.146 \left(\frac{m}{s^2}\right)^2$
6207	$1.236 \frac{m}{s^2}$	$1.624 \left(\frac{m}{s^2}\right)^2$

While for the 1687 RPM case, the bearing data follow a determined progression, the vibration RMS and Variance Features at the 1103 and 640 RPM are more intermixed as shown in Figure 4.27 (a) and (b), respectively. For the 640 RPM case, the bulk of the data appears to follow a similar progression as the 1687 case. However, a larger outlier group appears separate from the bulk of data

related to the 6205 case. The variation in the data stems from additional noise at the low-frequency level. The likely cause stems from an incorrect configuration in the test stand or a loose component. For the 1103 RPM case, the increase in the 6206 data possibly stems from a similar reason. The jump and variation between the different bearing sizes could be the result of the different bearing test locations.
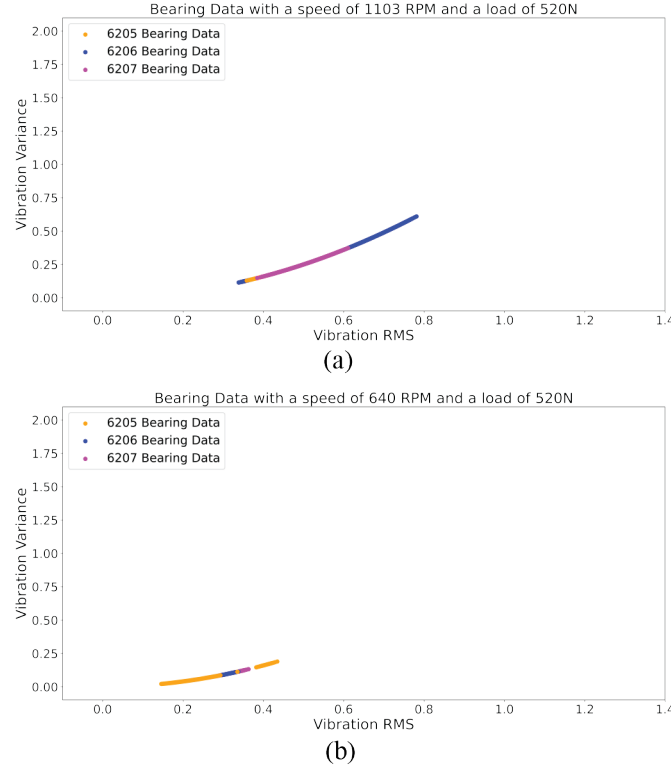
Figure 4.27: 6205, 6206, and 6207 bearings at the (a) 1103 RPM and (b) 640 RPM levels

In addition to the time series data, frequency representations were also considered to establish a bearing baseline between the different bearing stages. Figure 4.28 (a) shows the low-frequency (up to 1000 Hz) representation of the bearing vibration data with no filter applied. In addition to the bearing frequency data, the bearing frequencies are added, shown in black, and the speed and harmonic frequencies are shown in red. The speed frequency corresponds to the running speed of the equipment. The harmonics are the resulting impulses caused by initial equipment frequencies. These frequencies are the result of specific equipment events. The corresponding amplitudes are a representation of the FFT computation based on the absolute value of the complex numbers normalized based on the number of samples in the FFT calculation multiplied by two.

The "Bearing 5 6205" variation shows different harmonics and frequency peaks relative to the bearing data with a "K". The primary difference between the two bearings is the manufacturer and the location where the bearing data was collected. An interference from the surrounding equipment could have caused additional excitation of the speed frequencies for the bearing. The black peaks between 100 Hz and 150 Hz correspond to the bearing frequencies. The first black line closest to 0 Hz corresponds to the bearing BPFO. The second black line corresponds to the BSF, and the last black line is for the BPFI. The first two frequencies overlap with the bearing FFT amplitudes. The bearing 5x speed harmonic speed is close to the BPFI (within 20 Hz) and appears dominant over the BPFI.
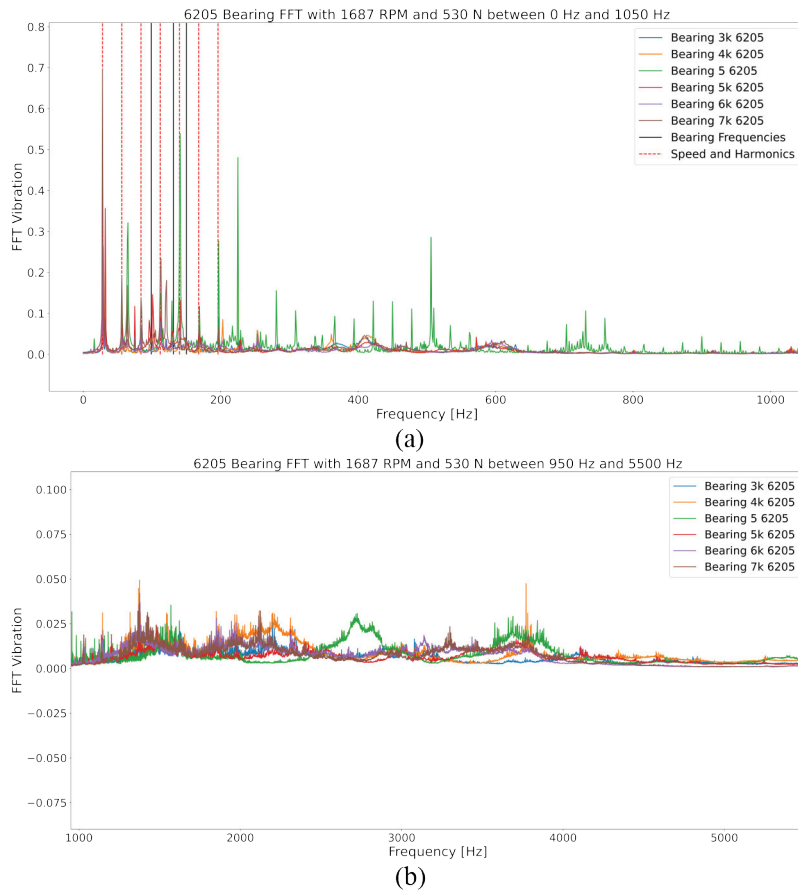
Figure 4.28: 6205 frequency representation for a low-frequency representation (a) and a high-frequency representation (b)

Figure 4.28 (b) shows the higher-frequency (between 1000 Hz and 5500 Hz) representation of the bearing vibration data. In the higher frequency data, there appears to be more variation from

case to case. The frequency peaks appear more distributed over a wider range than the sharp peaks in the lower frequency range. The distributed FFT peaks could correspond to additional vibration harmonics from the lower frequency data. Generally, the high-frequency peak data variation can change in a case-by-case scenario for the bearing. The lo- frequency peaks should be present in each bearing tested. An expected hypothesis examines the high-frequency variation to determine how they change relative to the baseline condition and the different failure modes tested.

Similar trends are found in the 640 RPM and 1103 RPM cases with respect to the bearing frequencies noticed and speed harmonics. Figure 4.29 considers the other speed cases in their frequency representation with respect to the speed frequencies and bearing characteristic frequencies. The FFT amplitudes appear to decrease with respect to the speed. In addition, the initial speed frequencies and harmonics change in correspondence with the bearing speed. The change is also reflected in the bearing frequencies as those decrease with respect to the bearing speed. The change in bearing frequencies is due to the dependence on the bearing speed. The variation between "Bearing 5" and the "K" bearings is still present in the data and is more noticeable in the 1103 RPM case (Figure 4.29). It should be noted that some of these variations could explain why the Bearing 5 6205 data appeared to differ from the other bearing data in Figure 4.19.

Figure 4.30 contains the average FFT response for the different bearing sizes at the 1687 RPM and the 530 N. A low-frequency representation between 0 and 1050 Hz (a) and a high-frequency representation between 950 and 5500 Hz (b) are provided to show the different amplitudes for each bearing at the corresponding frequencies. The 6207 bearing has the largest response from the speed harmonic frequencies, followed by the 6206 and 6205 bearings. Outside of the speed harmonics, the bearing frequencies also remain similar. However, the BPFI (the highest black line frequency) does not appear with respect to a particular peak in either the 6206 or 6207 data as it does with the 6205 data. In the higher frequency representation (Figure 4.30 (b)), the bearing data appears to have the most variation between the 1000 Hz and 4000 Hz range. The 6206 bearing appears to have a mixture of the frequency response between the 6205 and 6207 bearings. In the range of high-frequency data, there were overlapping peaks between each class, suggesting a level of similarity between the different bearings, despite the change in size

The similarity in the low- and high-frequency could come from the similarities in the bearing size. Table 4.18 contains the bearing frequencies for the 1687 RPM case for the tested bearings. Despite the variation in the bearing sizes, there are similar frequencies between each case. The
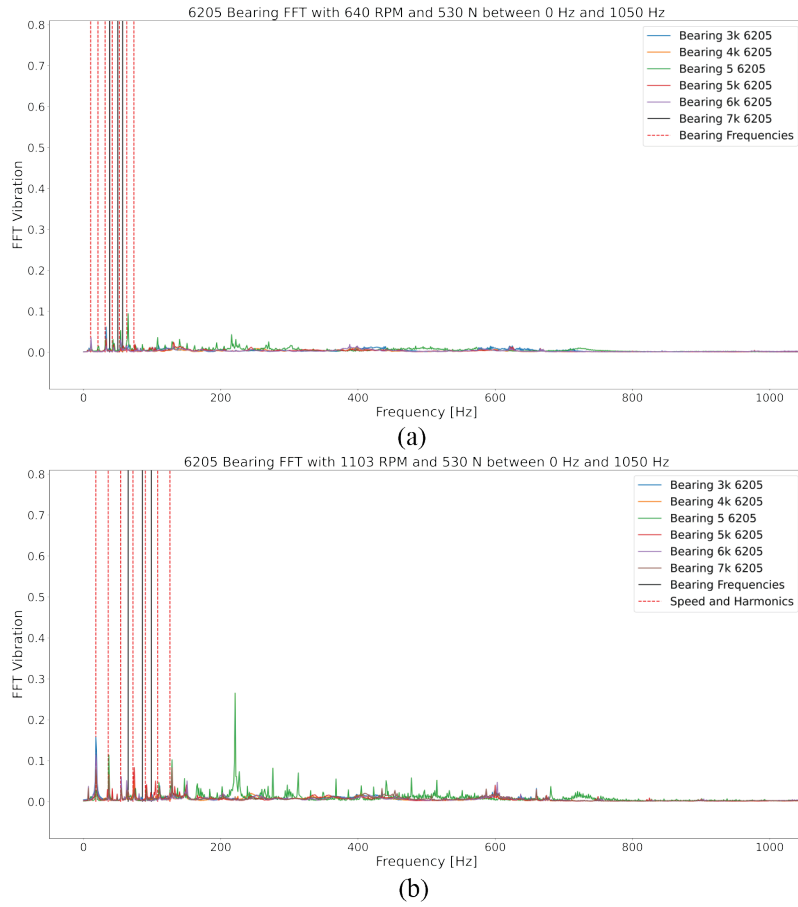
Figure 4.29: The average horizontal vibration FFT for the (a) 640 RPM and the (b) 1003 RPM cases for lower frequency data.

greatest deviation comes from the rolling element frequency; the largest is only 3 Hz. The difference could be accounted as normal during operation as the equipment attempts to maintain a set speed. Hence, the table points to why the frequencies between the different bearing sizes overlap.

Table 4.18: Bearing Defect Frequencies for 1687 RPM and the 6205, 6206, and 6207 cases

Bearing Size	Bearing Inner Race Frequency	Bearing Outer Race Frequency	Ball Spin Frequency	Fundamental Train Frequency
6205	151.17 Hz	100.08 Hz	131.62 Hz	11.12 Hz
6206	151.63 Hz	99.61 Hz	129.04 Hz	11.07 Hz
6207	151.71 Hz	99.53 Hz	128.61 Hz	11.05 Hz

In addition to the FFT, an additional time-frequency technique was considered, the Hilbert-Huang Transform. The HHT was considered for both the steady-state and varying condition data. Utilizing this method reduces the skewed information that occurs as a result of the shifting frequency

148
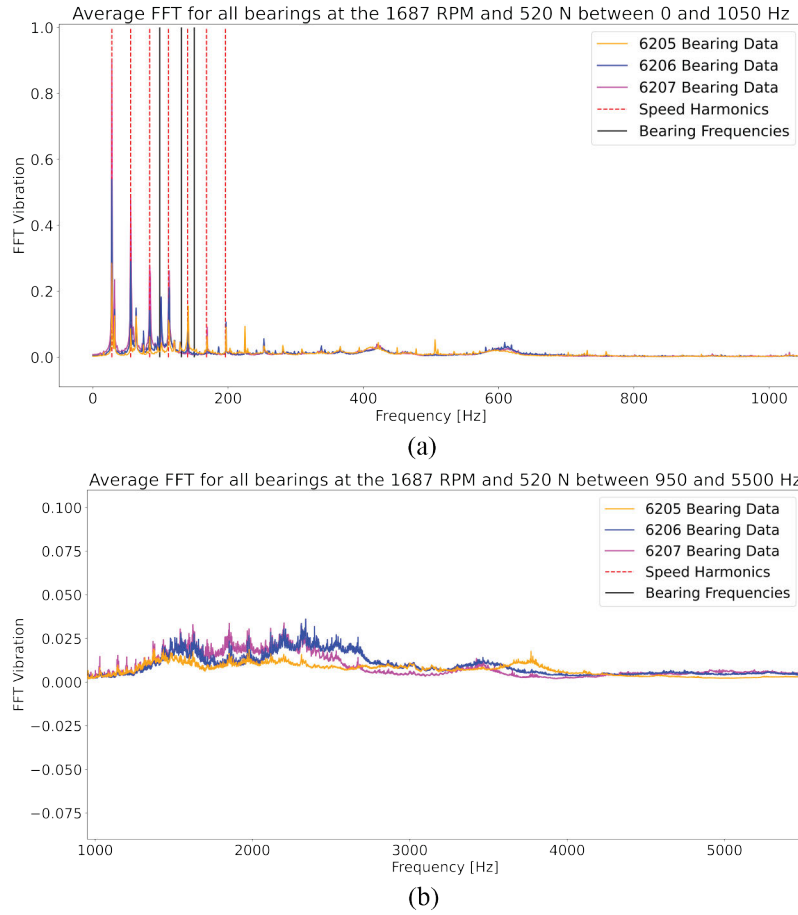
Figure 4.30: Average FFT for all bearings at 1687 RPM and 530N for the (a) low-frequency and (b) high-frequency cases

information seen with the FFT. The following library was used in Python to compute the Intrinsic Mode Functions (IMFs) and conduct the Hilbert spectral analysis, "emd" [499]. It was created using the work conducted by Huang *et al.* [488] in their initial formulation of the HHT. Furthermore, the results were checked using the "emd" function in MATLAB. Figure 4.31 and Figure 4.32 show the first 12 IMFs for the bearing response. For each IMF, there are differences in the response; however, as the bearing frequency decreases, the bearing IMF range also decreases. As the range decreases, the bearings' vibration response moves to an average line. In the 1103 RPM case, IMFs 7 and 8 deviate from the expected bearing vibration response, where Bearing 3K has more impulse events than the Bearing 4K and 6K data.

Figure 4.32 shows more variation between the different bearing cases in the IMF 1 case. However, as the individual IMFs are examined, there is a less run-to-run variation for each case.

149
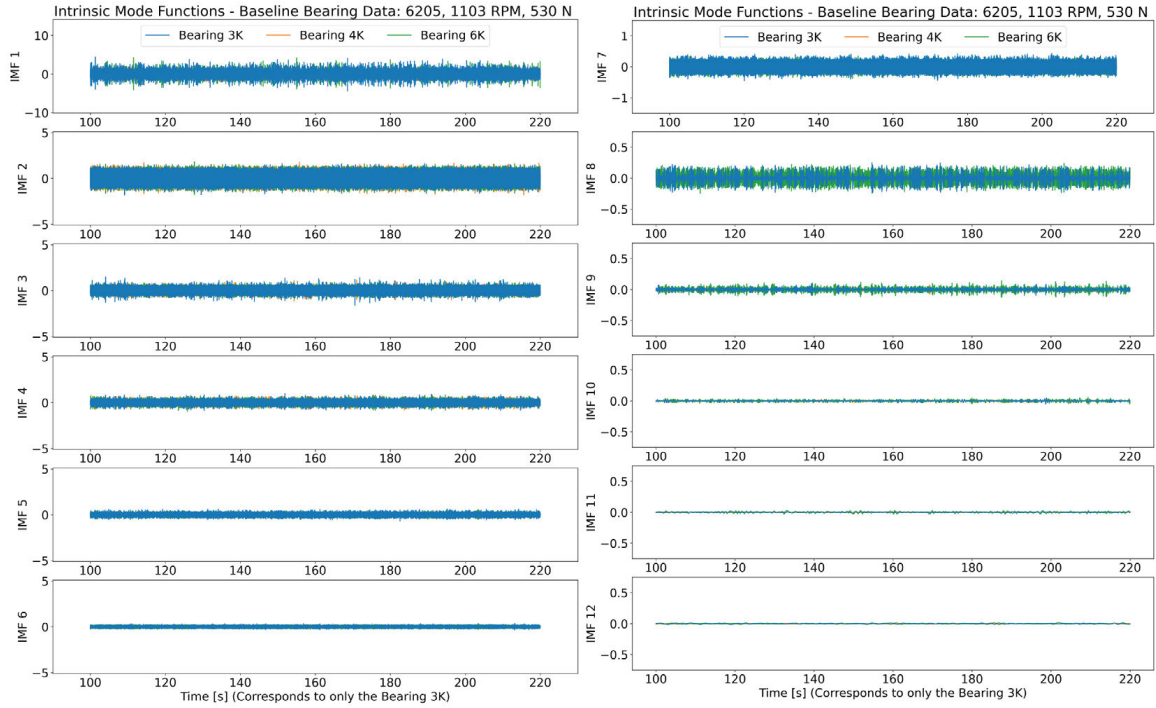
Figure 4.31: The IMFs for three 6205 bearings in 1103 RPM and 530 N

The IMF 2, 6, and 7 notice additional variation in each bearing case. These could correspond to additional frequency variations from the bearings due to the testing. In addition, the evaluation of these different IMFs helps isolate what frequency bands remain the same and which may be affected by additional noise from the system. In both cases for the 6205 bearings (Figure 4.31 and Figure 4.32), the IMF functions from 7 on appear to show the impulse events that may be more common at lower frequencies. The steady-state response in the IMFs before and imply that there is not as much variation in the run-to-run case for each bearing for high frequency responses.

Figure 4.33 and Figure 4.34 shows the IMF responses for select 6206 bearings at the 1103 and 1687 RPM cases, respectively. In the IMF 1 case for the 1103 RPM case, Bearing 3K shows variation during the testing phases. Bearing 3K 6206 provides an example of how the bearing response can change during testing as a result of adjustments made during the testing process for a particular bearing. In this instance, a looseness in the bearing housing affected the different bearing responses. In the first two IMFs, the bearing response appears to remain the same across the three different bearings, despite the variation in the IMF 1 case. In the IMF 3 case, there is an additional variation between 180 and the end of the bearing signal. That carries over until the IMF 7 case.
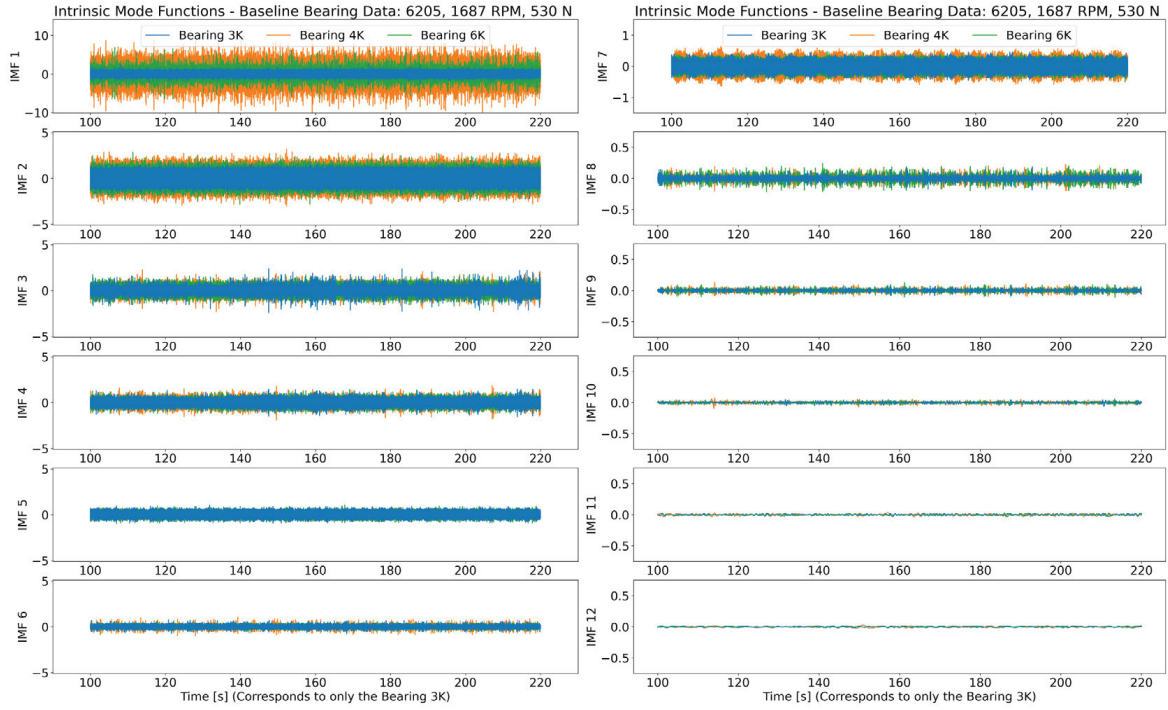
Figure 4.32: The IMFs for three 6205 bearings in 1687 RPM and 530 N

Tightening the bearing structure could provide a reason for the difference in bearing response. In the IMF 7 case until the end, the IMFs appear similar, suggesting that there are also similar frequency responses between the different cases. Overall, the bearing vibration may appear similar on a high level at the time series scale; however, when broken down into different componenets, the vibration amplitude is split between differnet frequencies and those frequencies may change during operation due to system or operating condition changes.

Figure 4.34 provides the expectation for bearing response on the steady-state case. However, in this instance, there is more variation in the high frequency response as shown in the IMF 3 as compared to the 6205 cases. In comparison to the IMFs from the 6205 case, the response appears similar for the bearing high frequency cases, even though the IMF 1 appears differently. However, in the lower frequency content, the bearing responses appear differently as compared to the 6205 and 1687 RPM cases. The difference in the impulse events could come from the differences in the bearing dynamics as a result of the bearing stiffness and deformation. The bearing force does not change as a result; however, the resulting deformation does change as a result of the bearing geometry, even under similar operating conditions. The separation of the different IMFs though help isolate the

151
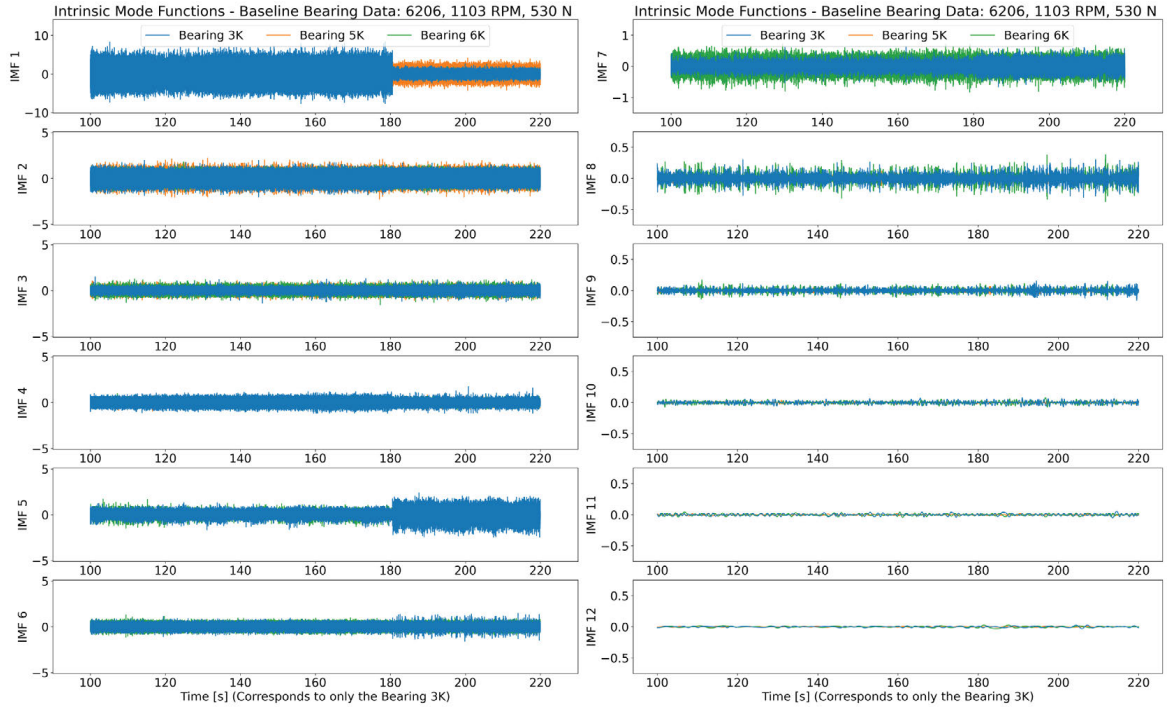
Figure 4.33: The IMFs for three 6206 bearings in 1103 RPM and 530 N

differences in the frequency responses between the different bearing sizes

Figure 4.35 and Figure 4.36 show the IMF functions as the speed changes during the bearing collection. Figure 4.35 considers the 6205 case, and Figure 4.36 considers the 6206 case. The first item to note is that not every response matches perfectly based on the testing process. The deviations could stem from the difference in testing procedure during the data collection. As an example, someone could have mistimed the decrease and increase in speed. Despite the lack of temporal overlap, it is clear from the IMF 1 case that there are increases and decreases with each collection. Incidentally, there is still a run-to-run variation between each bearing test.

In Figure 4.35, the IMF 1 case have a different appearance case-to-case. The initial high frequency does not appear to change as expected with only a slight change in the Bearing 3K and 4k. Bearing 6K does not appear to change However, the bearing signature has a similar appearance for IMFs 2 through 5 on a case-by-case basis that appears to match the bearing change in operation. In IMF 5 through 7, there is an additional increase in the Bearing 3K. That is due to an increase in the speed, due to the test taker increasing the speed, instead of decreasing the speed. The slight increase does not appear to translate into the earlier IMF functions. The difference in amplitudes
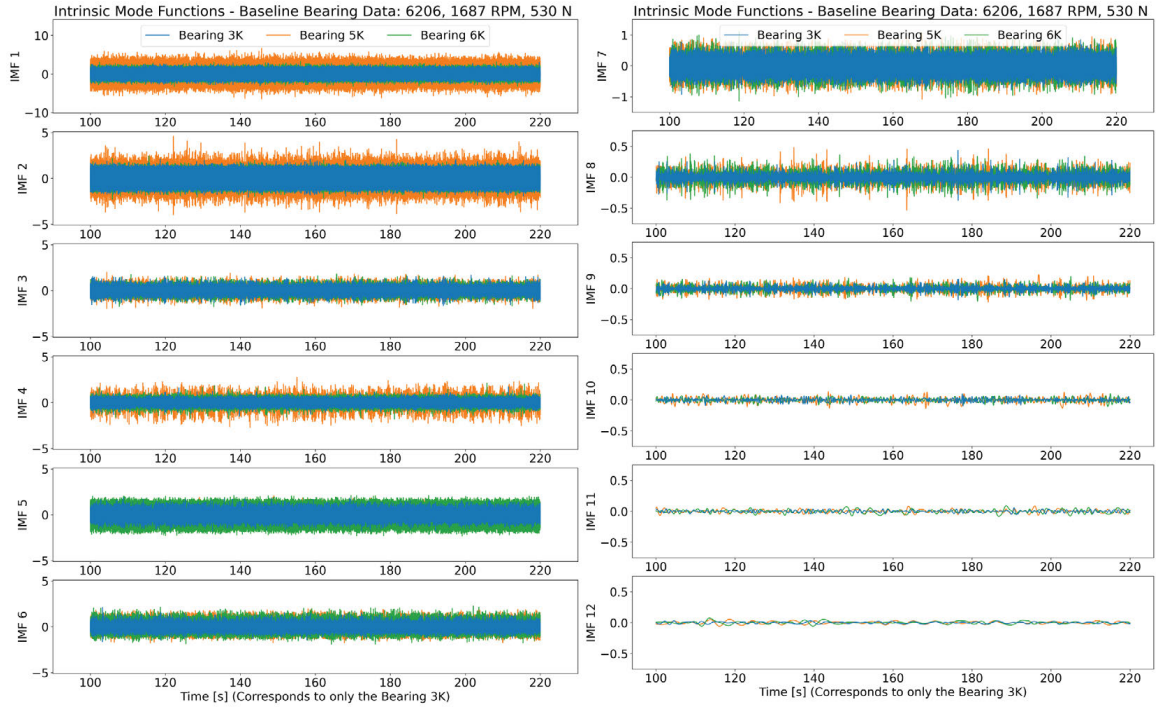
Figure 4.34: The IMFs for three 6205 bearings in 1687 RPM and 530 N

could corroborate The impulse events that appear in IMF 8 to IMF 11 seem to remain similar to the steady-state considerations.

In Figure 4.36, each bearing appears to have a different response. Bearing 5K appears to have more impulse events that carry over in the high frequency range. IMF 2 has significantly higher amplitudes in the Bearing 5K signatures over the other bearing signatures. Despite the change in bearing amplitude, the drop and increase in amplitude does match the expected response as a result of the operating conditions. The larger amplitude remains for Bearing 3K in the case of IMF 2 and IMF 3. However, starting with IMF 2 and IMF 3, the operating condition variation comes through with the Bearing 5K and 6K. The change in the signature for IMFs 2 through 4 matches the data collection procedure for the bearing speed varying case. In contrast to bearing 6205, the speed-varying signatures continue in the IMF 8. The impulse events continue in the low frequency responses, like the steady-state speed considerations in Figure 4.34.

The bearing signatures appear to vary widely on a case-to-case basis for the 6205 and 6206 cases. The primary reason that could contribute to the different responses stems from the sifting process [500]. The sift runs based on the local minima and maxima of a signal to meet the two

153
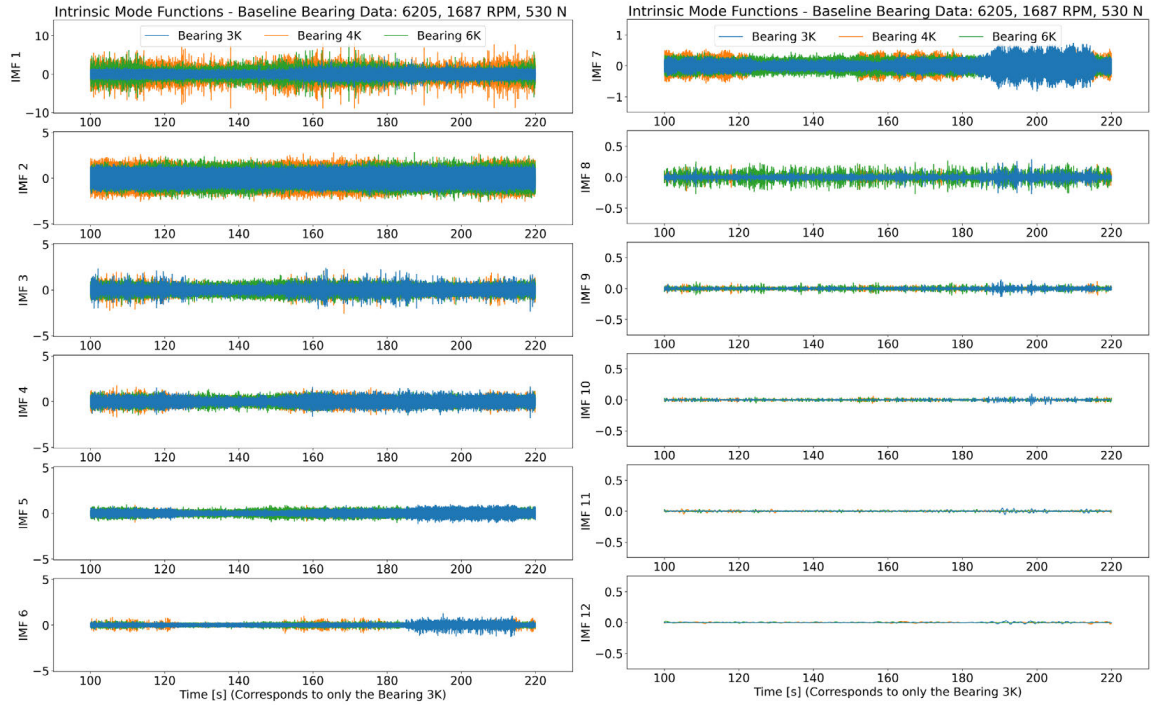
Figure 4.35: The IMFs for three 6205 bearings in 1687 RPM and 530 N under speed varying conditions

expected IMF criteria. The sifting ends for the IMF extraction once these criteria are satisfied. However, that leads to a different number of IMFs extracted from the bearing. For example, in the steady state case for the 6205 bearing at 1687 RPM Bearing 3K had 19 IMFs extracted, Bearing 5K had 17, and Bearing 6K had 18. The difference in the IMF amplitudes may be explainable based on the variation in the total number of IMFs. Hence, utilizing the spectral analysis of the Hilbert analysis, specific frequency bands are isolated for each IMF.

For a brief introduction, Hilbert spectral analysis is conducted by applying the Hilbert transform to a real signal. The Hilbert Transform differs from the Fourier Transform by shifting the sign of the frequency components by a multiplication factor. The shift occurs based on a positive 90 degrees for the negative frequency components and a negative 90 degrees for the positive frequency components. A double application of the Hilbert transform would return a negation of the original transform. In HHT, the Hilbert spectral analysis extracts the instantaneous frequencies and amplitudes for the vibration signal.

Figure 4.37 shows the spectral signature for the steady state cases for the Bearing 5K (a, c, e) and Bearing 6K (b, d, f) for IMF 1 (a, b), IMF 5 (c, d), and IMF 6 (e, f). Despite the difference in

154
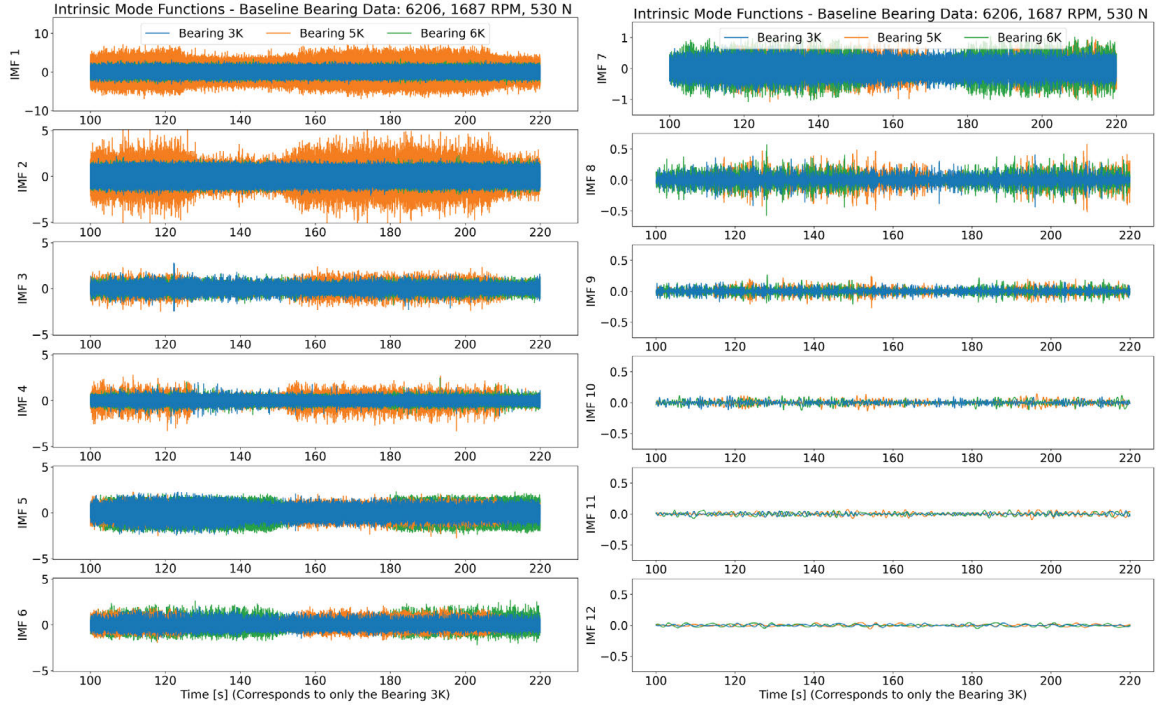
Figure 4.36: The IMFs for three 6206 bearings in 1687 RPM and 530 N under speed varying conditions

amplitudes in the raw IMF functions, it would appear that the bearing frequency ranges are similar. Hence, the variation in amplitudes could stem from the bearing tested at that particular instance. The bearing variation is common in other datasets. For example, in the IMS bearing dataset [408], the bearings baseline signatures change between Test 1 and Test 2 and 3. The variation could have been run-to-run in the IMS bearing dataset. The similarity in the IMF frequency signatures leads to the conclusion that the IMF signatures are extracted at a similar iteration. The time refers to the duration of the vibration time length in seconds.

However, despite the similar frequency bands, the dominant frequencies are different in each bearing case. For example, the frequency amplitudes in Figure 4.37 (c) are centered around the bearing at 100 Hz. In Figure 4.37 (d), the frequency amplitudes are around 60 Hz. However, frequencies are still captured in each frequency band encompassing the 60Hz and 100 Hz ranges, respectively for either bearing. They do not represent the dominant frequency, though. In the IMF 1 and IMF 6 frequency ranges, the response are more aligned. The resemblance could be related to commonalities in terms of the operating conditions. The importance of the difference for the IMF 5 signatures is that these frequencies encompass the expected bearing characteristic frequencies.
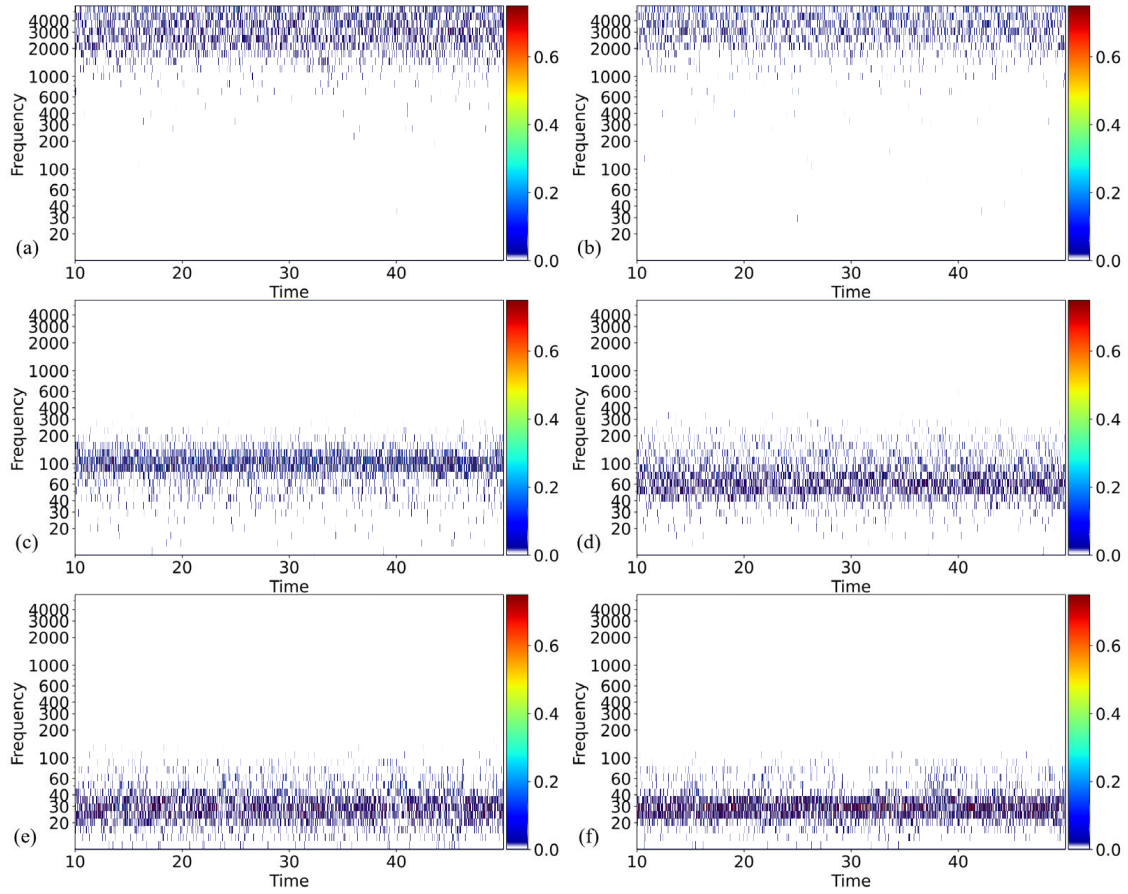
Figure 4.37: Hilbert spectral analysis for Bearing 5K and Bearing 6K under the 1687 RPM and 530 N cases for the steady-state case. (a) Bearing 5K IMF 1, (b) Bearing 6K IMF 1, (c) Bearing 5K IMF 5, and (d) Bearing 6K IMF 5, (e) Bearing 6K IMF 6, and (f) Bearing 6K IMF 6

Hence, in later testing for failure modes, the frequencies in this IMF range are expected to change relative to the bearing failure mode.

Figure 4.38 contains the bearing HHT analysis for the 6206 bearing case for the Bearing 5K (a, c, e) and Bearing 6K (b, d, f) at IMF 1 (a, b), IMF 5 (c, d), and IMF 6 (e, f). In IMF 1, there is no noticeable change in the bearing signature compared to the IMF range in the steady state case from Figure 4.37. In the Bearing 5K case, there is a lower response before the 30s for the amplitudes in the Bearing 5K IMFs 5 and 6. Hence, it is easy to track and notice what frequencies change as the operating conditions change. For the Bearing 6K case, the IMFs do not appear to change relative to the time. This highlights why there are variations in the IMFs between the different bearings due to the testing procedure in lining them up based on time. It does show that the base speed in the speed-varying case does match a similar speed in the steady-state cases. The load-varying cases

156
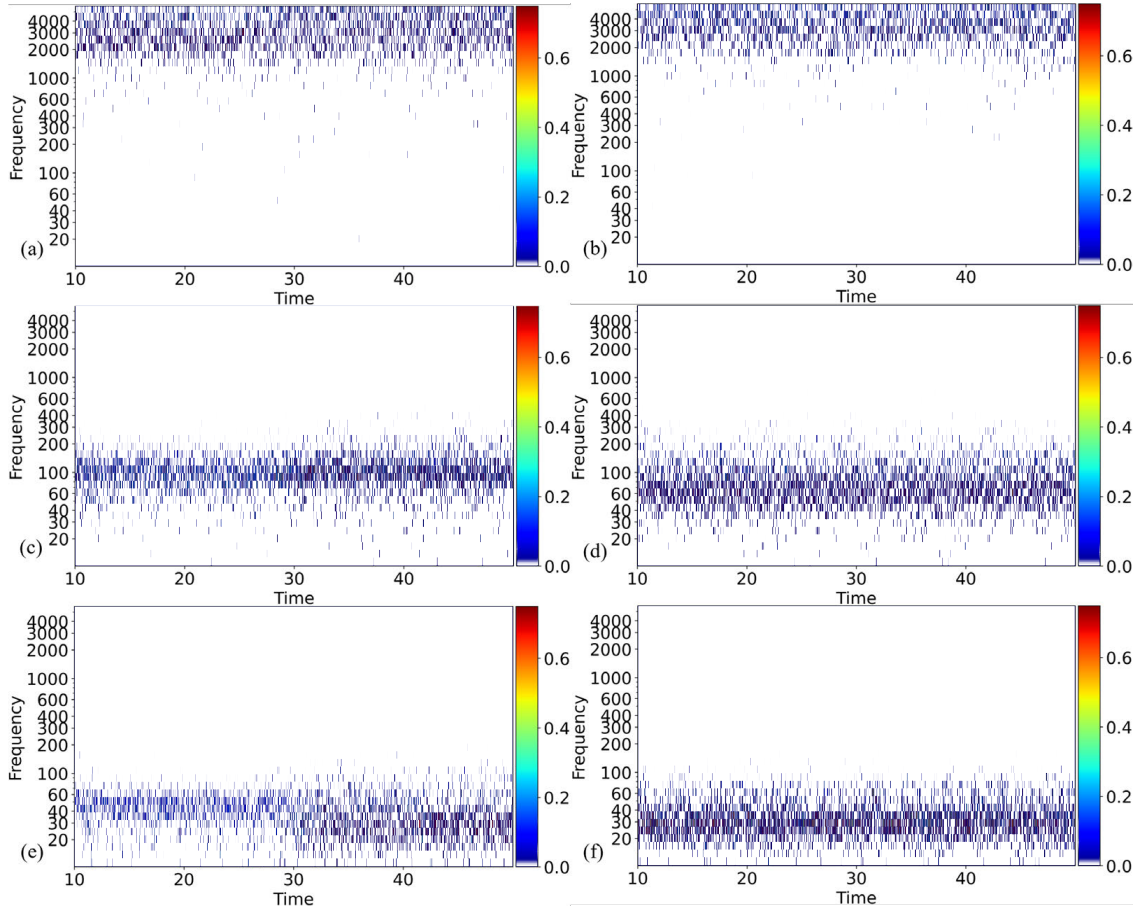
Figure 4.38: Hilbert spectral analysis for Bearing 5K and Bearing 6K under the 1687 RPM and 530 N cases for the speed-varying case. (a) Bearing 5K IMF 1, (b) Bearing 6K IMF 1, (c) Bearing 5K IMF 5, and (d) Bearing 6K IMF 5, (e) Bearing 6K IMF 6, and (f) Bearing 6K IMF 6

were not considered for the baseline characterization due to the lack of variation in the baseline signatures for the statistical measures. However, these are further investigated for the defect cases if they are more readily apparent.

The purpose of this section was to characterize and discuss the baseline for the bearing test system. In this manner, it is possible to identify variations for each defect stage tested for the bearings. Experimentally, three different analysis methods were considered to establish the bearing expected response: statistical features, Fourier analysis, and HHT. Each analysis method is considered to consider different aspects of the bearing signature. In Chapter 5.1.1, the baseline response is compared to the physical analysis and forms the validation for the physics-based model.

4.3.2 Characterization in Failure Conditions

Following the baseline data generation, the failure modes are further examined in terms of the types of damage, the damage appearance, and the resulting failure signatures in the failure data. At this stage, the primary examination is identifying the differences between the baseline and fatigue cases. These conditions are investigated using the same analysis methods as described in the bearing analysis section and for verifying the baseline conditions. Furthermore, supporting images are added to demonstrate the changes experienced in the bearings due to the different damage methods.

4.3.2.1 Comparison of Bearing Fatigue

Figure 4.39 contains the initial defects for the bearing fatigue cases on the inner race for each bearing size. The initial defect case was meant to have a 1 mm diameter in each case. Based on the tool size and the radius, the expected depth for each defect was 0.1 mm. The defect design was based on prior literature related to inducing bearing defects. In addition, defects for ball bearing elements are generally designed as a point defect. Appendix D contains the Solidworks tool representation for points of contact in Figure D.2. From Figure 4.39, the bearing defect case had average size of 1.16mm based on the measurement lines attached to each figure. Each line represents approximately 0.254 mm, or a hundredth of an inch. The bearing defects are centered in the bearing raceway by touching off either side of the bearing with the tool and then dividing by two. par

The dashed lines show the edge for each bearing defect measurement based on the expected diameter. However, there were some errors in the bearing measurement, as shown in Figure 4.39 (a) and Figure 4.39 (b). In Figure 4.39 (a), the bearing defect appears to have a bump in the defect placement. In Figure 4.39 (b), there is a noticeable bump next to the right dashed line in Figure 4.39 (b). In Figure 4.39 (c), the defect is slightly oval. A possible reason for the not-near-perfect appearance of each bearing defect is that the tool "walked" along the bearing surface. The probable reason is the hardness of the material relative to the force applied by the CNC machine. The tool could not maintain the stiffness necessary to remain in place. Hence, the deviations in the bearing surface defect for each case. It is possible that these would affect the experimental results of contrived fatigue damage; however, compared to real-world defects, defect shapes would appear non-perfect. Hence, it is considered within the acceptable limits for the defect. Future considerations could involve using a stiffer tool or machine to prevent the defect from changing on the surface.
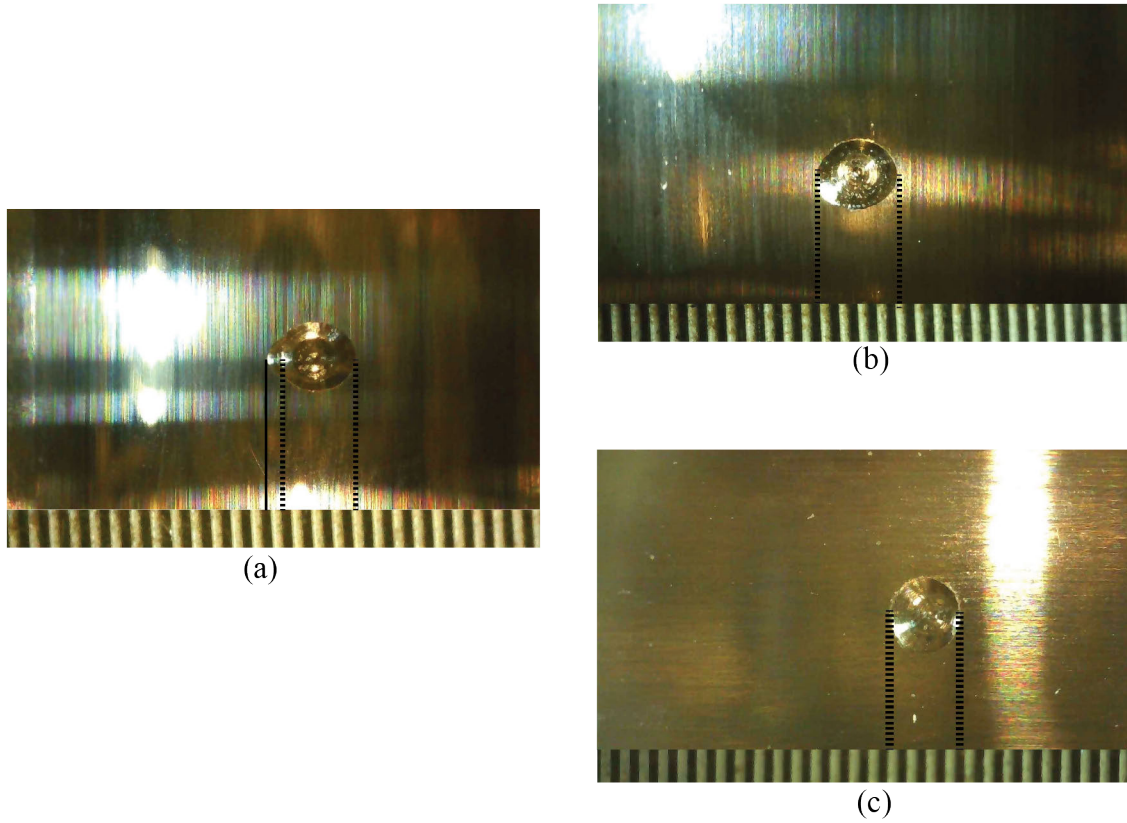
Figure 4.39: Example of an initial defect on the inner race for (a) 6205 case, (b) 6206 case, and (c) 6207 case.

Figure 4.40 considers the 3 mm cases for the 6206 and 6207 cases, respectively. In this instance, the cuts are not nearly as smooth as the initial defects in the bearing fatigue case shown in Figure 4.39. The first item to note is that neither of these defects reaches the 3mm mark before testing. They do represent an increase in damage; however, the tool used began to wear down during the damage generation phase. The tool damage was noticed when creating the 7 mm case. The damage increase was selected to exponentially increase the bearing damage width. The average damage from the 6206 case is in fact a 1.5 mm case and for the 6207 case, the damage measurement is 2 mm for the 6207 case. Hence, while the damage does not meet the proposed case, damage accumulation is present.

Figure 4.41 shows the initial case after the bearing was run for three hours at different speeds. There is not a noticeable change in the defect size along the edge. However, there does appear to be a greater discoloration along the base of the defect. The discoloration is related to the over rolling of the main defect. There are also a number of pock marks along the surface of

(a)   (b)
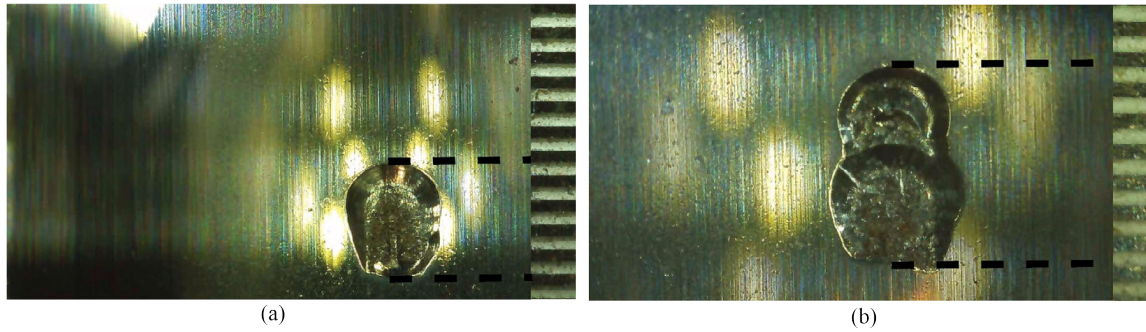
Figure 4.40: Bearing 3mm fatigue cases for the (a) 6206 case and the (b) 6207 case.

the bearing raceway around the bearing inner race defect. These could have occurred from the over rolling of the bearing defect particles from the CNC machine. The pot marks were not expected from the experiment; however, they do occur in real applications as the result of surface from damaged bearings.

Figure 4.41: Inner race defect case for the 6207 case after the experimental phase

Figure 4.42 shows the comparison of the different inner race defect cases for each of the speed cases. The different markers represent the speed cases for each bearing. The green represents the baseline case, and the red represents the failure case. For the 1687 RPM case, the bearing damage is separable from the bearing baseline case. However, in the 1103 RPM case, there is an overlap in the features between the damage and baseline cases. It is inconclusive based on the time

series analysis conducted what could be the difference between these two cases. In the 640 RPM case, the bearing defect case is higher than the baseline case.
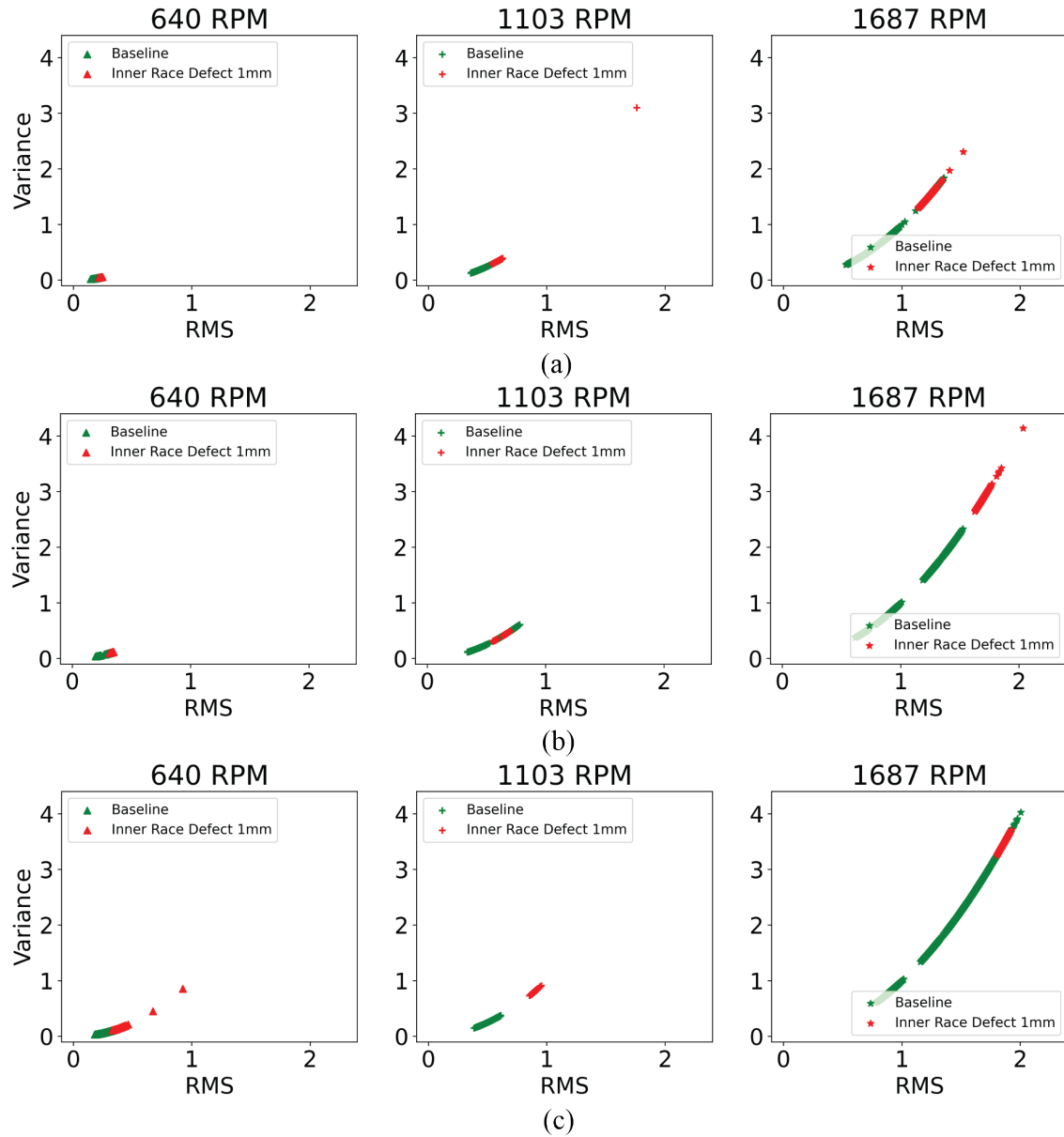
Figure 4.42: RMS and Variance comparison between the baseline and IR defect for the 1 mm case comparison for the different speed cases at (a) 6205 case, (b) 6206 case, and (c) 6207 case.

The resulting higher vibration is expected based on the defect overrolling and contact deformation. However, the overlap between the baseline and inner race defect cases is concerning, as the lack of separation could reduce the failure diagnosis. As an additional check of the time

161

features, Figure 4.43 shows the Kurtosis and skewness comparison between the baseline and 1 mm inner race defect. Based on observation, there is a significant amount of overlap between the inner race defect and the baseline cases between the 6205 640 RPM case, the 6205 1103 RPM case, and the 6207 640 RPM case. A level of overlap is expected in the lower speed cases, as the vibration excitation is not as separable as seen between the different defect cases. However, in the 6205 1103 RPM case, the similarity between the inner race defect and the baseline case was unexpected. From the test notes, and the time-based observation of the data, it is unclear what would have caused this overlap. Furthermore, this comparison only considers a partial data account of the baseline data. Several bearings in the 6205 and 6207 cases caused anomalous readings on the examination of the kurtosis and skewness. Appendix C further discusses the faulty baseline data and the effects of separation between the bearing features. Hence, the frequency analysis is necessary to corroborate and understand where the possible overlap in the data could occur.

Figure 4.44 shows the average bearing FFT for each bearing size of the baseline and IR defect case at the 640 RPM speed case. The bearing characteristic frequencies are added in at certain speed harmonics. In the figure comparison, all of the bearing data were used to create the average frequency response. For each bearing case, the low frequency response is shown to characterize the change to the bearing characteristic frequencies and their harmonics (Figure 4.44 (a, c, e)), whereas Figure 4.44 (b, d, f) shows the bearing characteristic frequency in relation to the bearing high frequency components (between 1000 to 10000 Hz).

For the bearings at the 640 RPM case, the frequency responses look similar in the 6205 and 6206 cases. Despite the overall similarity, there are shifting peaks within the bearing data that could correspond to the added bearing defect. In examining the different bearing characteristic frequencies, the first-order characteristic frequencies do not appear to see a significant spike in the data relative to the bearing baseline data. Additional peaks start to form around several of the harmonic order characteristic frequencies. In the 6207 case, larger differences between the baseline and IR defect cases are based on the shift in peak frequencies and amplitudes. For the bearing characteristic frequencies, the 10th order harmonic frequencies appear to have the most change in defect frequencies.

Figure 4.45 considers the 1103 RPM speed comparison between the baseline and IR defect cases. The 1103 RPM case for bearing 6205 appears to have the highest low frequency response. It is possible that the critical frequency was hit for the small bearing in relation to the large frequency.

162
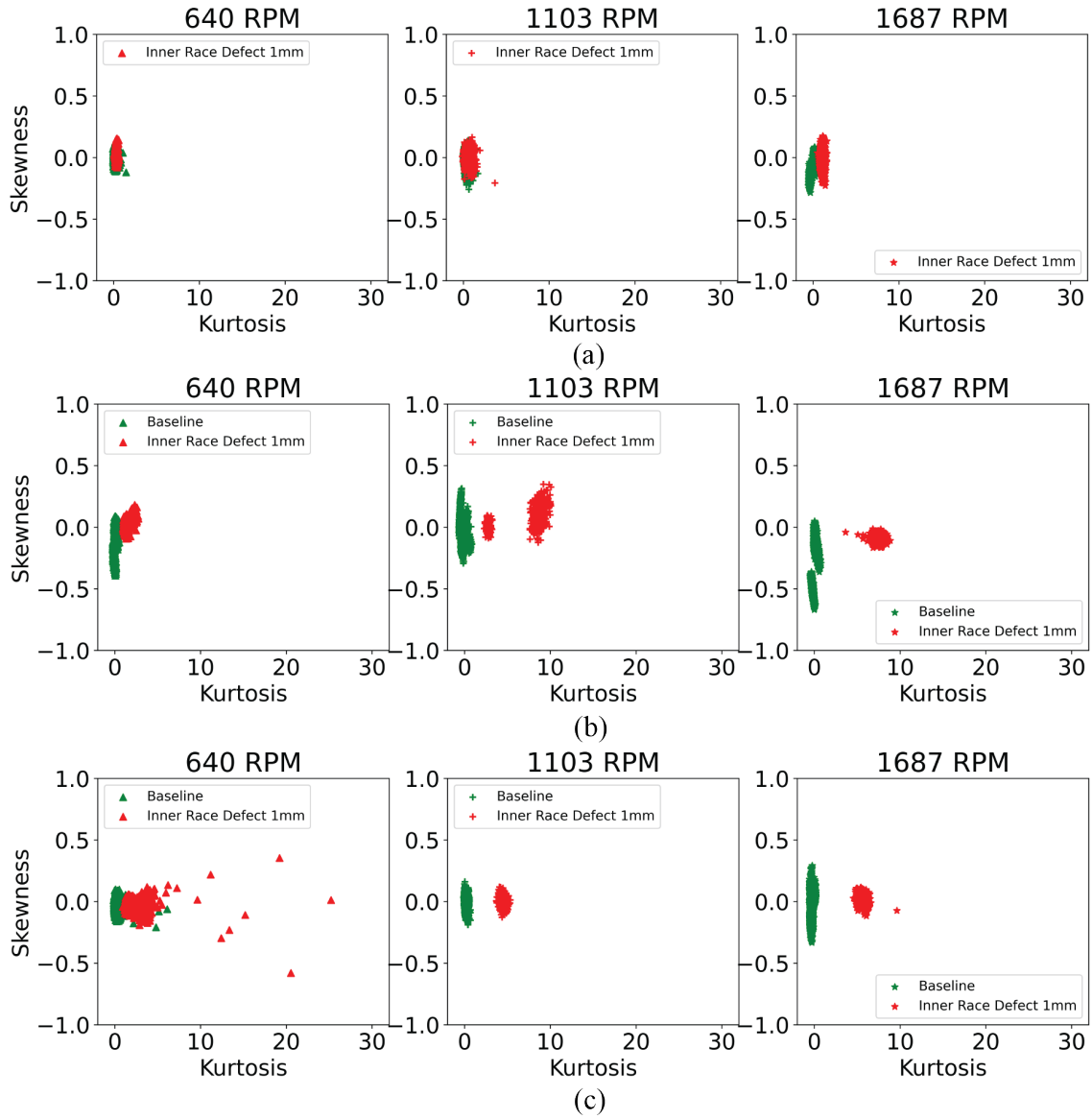
Figure 4.43: Skewness and Kurtosis comparison between the baseline and IR defect for the 1 mm case comparison for the different speed cases at (a) 6205 case, (b) 6206 case, and (c) 6207 case.

Additionally, it could signify additional looseness for that particular bearing in relation to the overall system. In the 6206 case, the baseline low frequency peaks are higher than the defect case. Additional peaks captured in the higher frequency data could signify the separability between the bearing baseline case seen in the time series data. For the 6207 bearing data, the bearing frequencies appear higher after the 1000Hz range and continue into the high frequency range. This could correspond to the increase in speed identified by the bearing data.
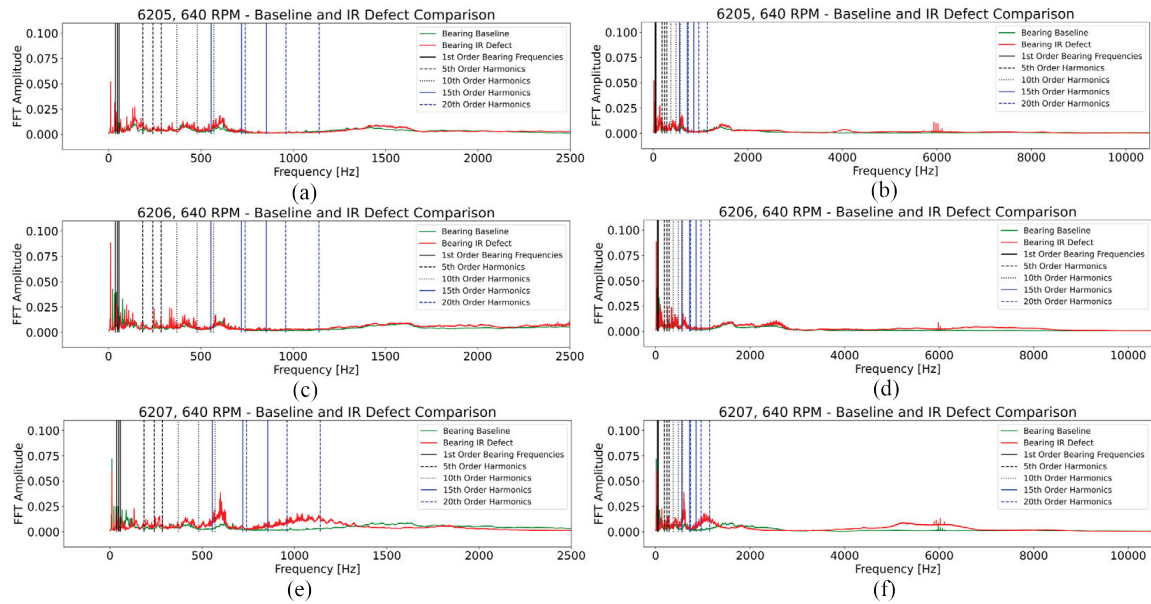
Figure 4.44: FFT representation for IR defects and baseline comparison at 640 RPM for the 6205 (a) 0 Hz - 2500 Hz range and (b)0 Hz - 10000 Hz, 6206 (c) 0 Hz - 2500 Hz and (d) 0 Hz - 10000 Hz, and 6207 (e) 0 Hz - 2500 Hz and (f) 0 Hz - 10000 Hz bearings
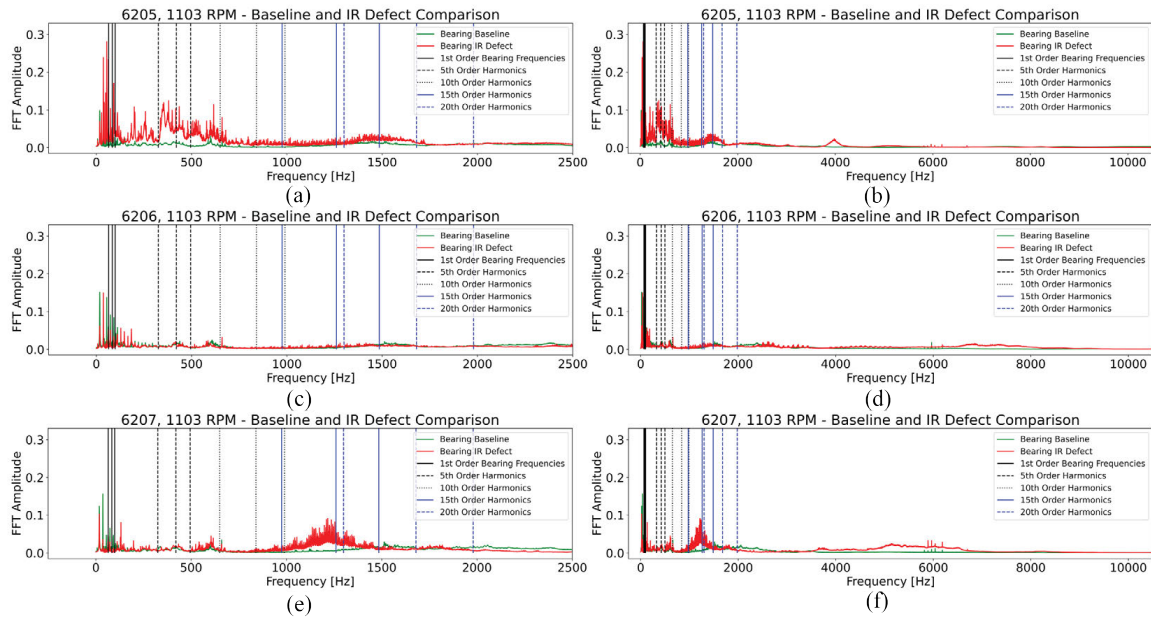
Figure 4.45: FFT representation for IR defects and baseline comparison at 1103 RPM for the 6205 (a) 0 Hz - 2500 Hz range and (b)0 Hz - 10000 Hz, 6206 (c) 0 Hz - 2500 Hz and (d) 0 Hz - 10000 Hz, and 6207 (e) 0 Hz - 2500 Hz and (f) 0 Hz - 10000 Hz bearings

Figure 4.46 considered the average FFTs for the 1687 RPM cases at the different bearing sizes. Similarly, for the 1103 RPM case, the 6205 bearing size had higher peaks at lower frequencies, whereas the low frequency peaks were lower at the 6206 and 6207 bearing sizes. In addition, the bearing response in the 6206 case appeared to spike in the higher frequency ranges, more so than in the 6207 size. Based on these additional observations, it is possible that the bearing fixture played a role in the different frequency responses. While the overall design of the two different fixtures is the same, the difference in mass could explain the difference in the bearing response. Consider a piece of string, fixed at one point and attached to a mass pendulum. The frequency increases if the mass decreases relative to everything else remaining the same. Utilizing second order system dynamics, the mass, spring and damping components could describe the change in frequency for the different bearing cases.
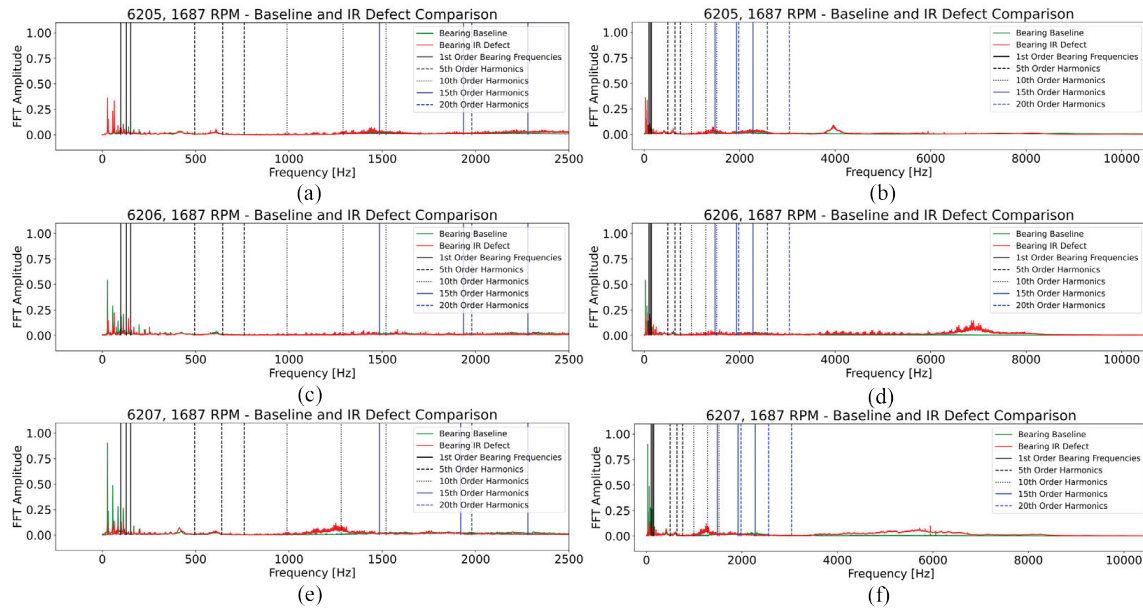
Figure 4.46: FFT representation for IR defects and baseline comparison at 1687 RPM for the 6205 (a) 0 Hz - 2500 Hz range and (b)0 Hz - 10000 Hz, 6206 (c) 0 Hz - 2500 Hz and (d) 0 Hz - 10000 Hz, and 6207 (e) 0 Hz - 2500 Hz and (f) 0 Hz - 10000 Hz bearings

Figure 4.47 compares the bearing baseline and defect case. The color bars are normalized to determine where the representations differ between cases. The first noticeable aspect is a lower representation of the low-frequency data in the baseline case compared to the inner race defect case. In addition, the inner race defect case appears to show impulse events happening periodically that corresponds to the defect introduction. Based on the frequency range, it is possible that the impulse

events could correspond to the speed.
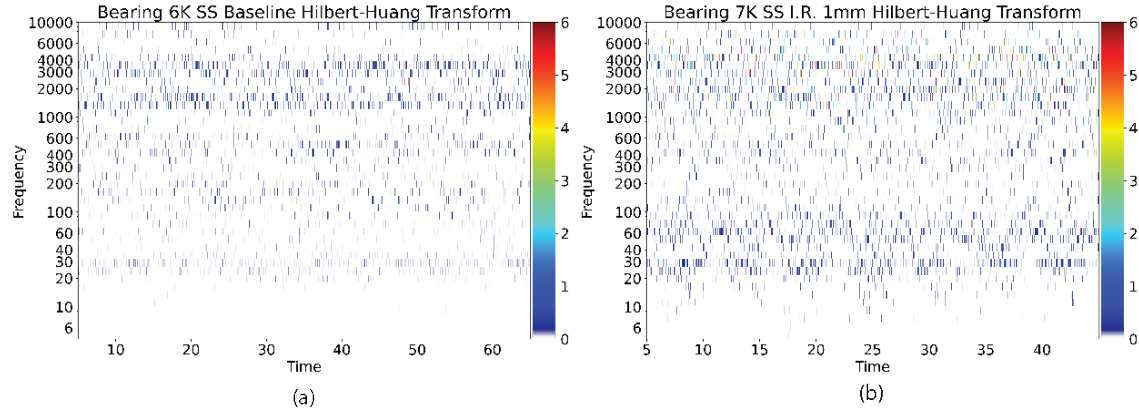
Figure 4.47: HHT representation of the 6205 bearing and 1687 RPM case for the steady state collections for the (a) baseline and (b) inner race defect case

Figure 4.48 shows the corresponding IMF bands for the inner race defect and a random bearing baseline case. The different IMF bands correspond to different frequency ranges as was discussed in the baseline characterization. For IMF cases 1 through 5, the inner race defect appears to have similar responses with different amplitudes. However, in IMF 6 and 7, the inner race defect case appears as an impulse event intermittently corresponding to the inner race defect case. It is possible to consider further that this shows the direct correlation between the defect introduction and the changes in vibration. The consistent impulse appearance represents the consistency of the rolling element over the inner race defect.

Figure 4.49 considers the 6206 bearing size for the 1687 RPM case in the frequency space. In the comparison between the bearing baseline and defect cases, the amount of low frequency signals is higher in the baseline case versus the inner race defect case. There are impulse defect cases in the 6206 case that are similar to what was examined in the 6205 case. A potential reason could stem from the test stand configuration and the introduction of a noise element that may have caused an increase in the bearing vibration for the baseline case or a reduced response for the inner race case. For example, a loose connection between structural elements could increase the noise in the baseline case, whereas as a tighter connection in the inner race case could dampen the expected amplitude. The high frequency amplitudes are higher in the inner race defect case than in the baseline case, supporting the differention between the baseline and inner race defect case. A similar trend was noticed in the 6205 bearing data as well.
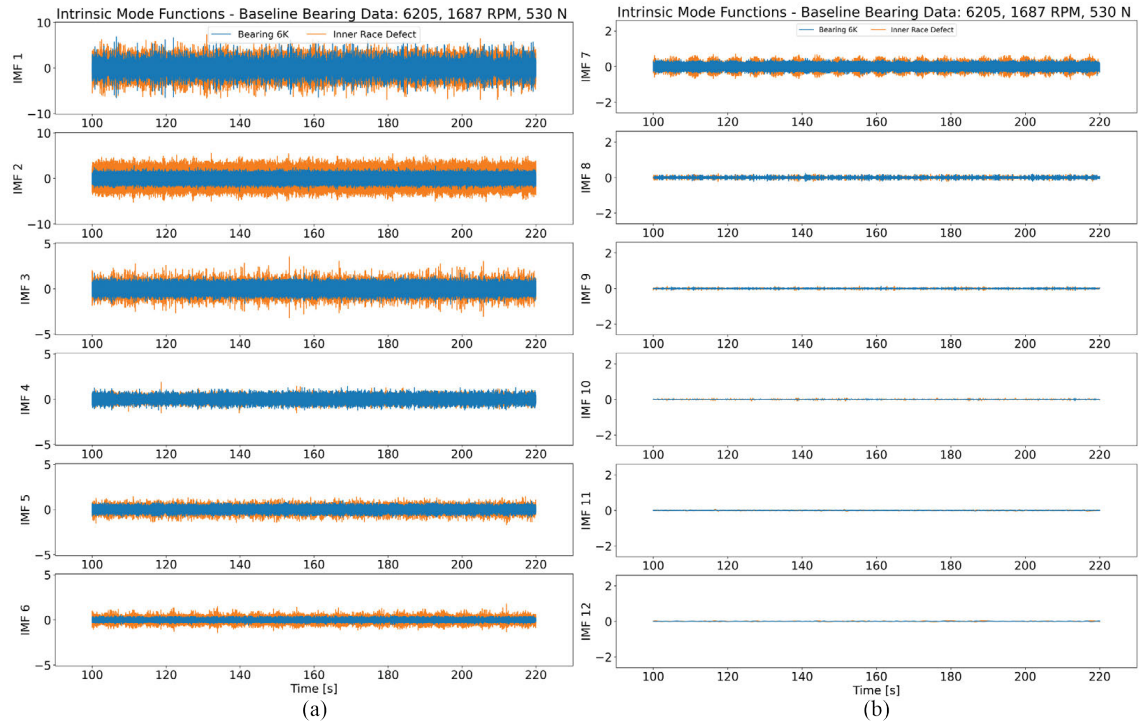
Figure 4.48: IMF representation of the 6205 bearings and 1687 RPM case for the steady state collections for the (a) baseline and (b) inner race defect case
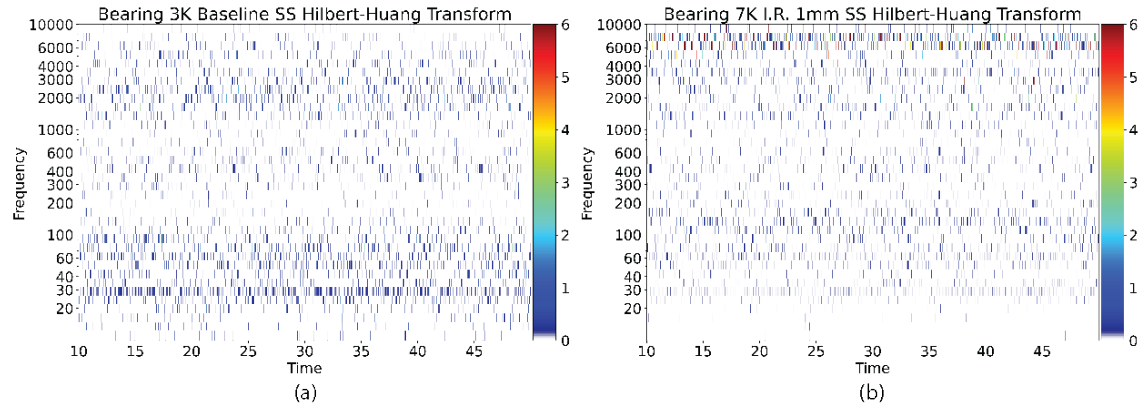
Figure 4.49: HHT representation of the 6206 bearing and 1687 RPM case for the steady state collections for the (a) baseline and (b) inner race defect case

Figure 4.50 considers the 6206 bearing size for the 1687 RPM case in the IMF representation and further corroborates the readings. In IMF 1 through 3, the readings show that the inner race defect has a higher amplitude as the vibration is recorded. The difference in the bearing IMFs supports the differences in the HHT response seen in that range. Similarly, the IMFs for cases 4

through 12 show that the bearing response in the baseline case is higher than in the inner race defect case. In Appendix D.3, an additional case is shown for a different baseline bearing to demonstrate the difference between the damage and baseline case and highlight the overlap between them.
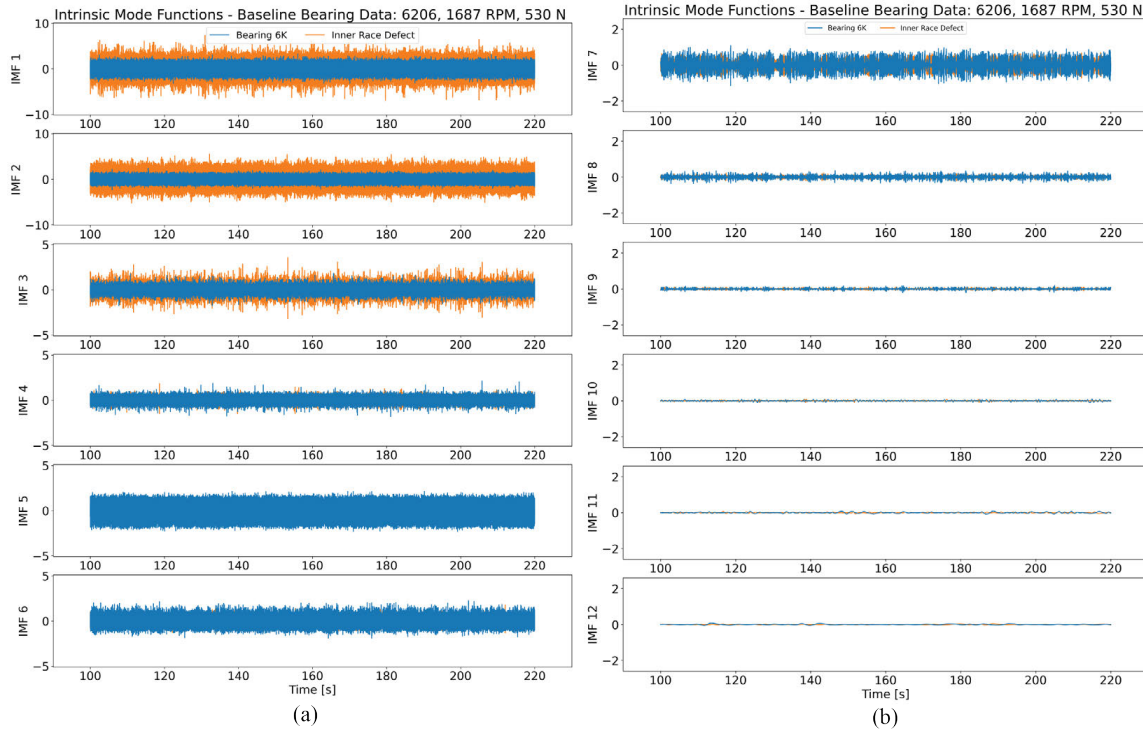
Figure 4.50: IMF representation of the 6206 bearings and 1687 RPM case for the steady state collections for the (a) baseline and (b) inner race defect case

Figure 4.49 considers the 6207 bearing size for the 1687 RPM case in the HHT representation. In the 6207 case, a similar trend is found in comparison to the 6205 case. In the baseline case for the Bearing 6k provided, a lower amount of impulse events appear relative to the defect cases. In the inner race defect case, the impulse events appear in the bearing case, similar to those in the 6205 case. The impulses have at a similar rate as shown by the expected frequencies when considering the characteristic frequencies in the baseline case. The high-frequency data appears to have similar components as shown with the 6205 and 6206 cases. The high-frequency signals correspond to the wear introduced by the bearing defect. The likely cause stems from accumulating the harmonic events found in the low-frequency data.

Figure 4.52 considers the 6207 bearing size for the 1687 RPM case in the IMF representation. The 6207 follows the same trend as the 6205 and 6206 cases for the high-frequency components
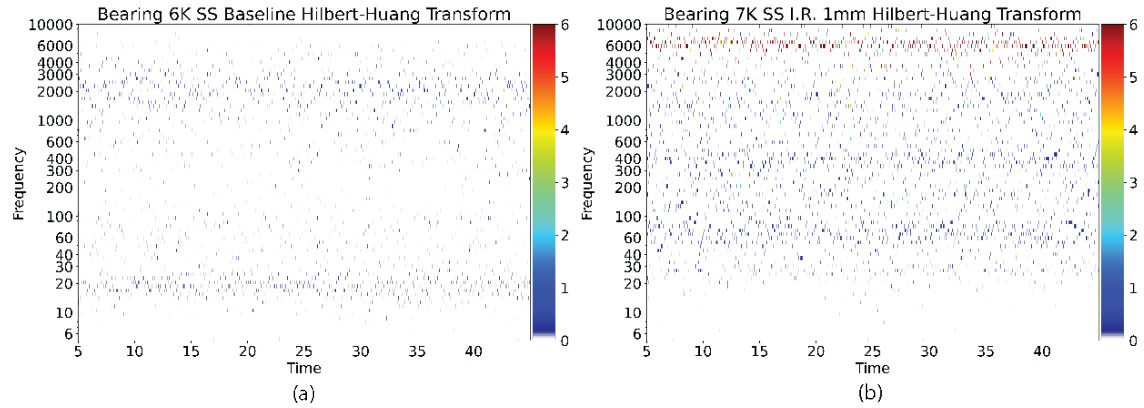
Figure 4.51: HHT representation of the 6207 bearing and 1687 RPM case for the steady state collections for the (a) baseline and (b) inner race defect case

between IMF 1 through 3. In IMF 4, it appears that the transient events begin to appear in this frequency range. However, as found in IMF 6 and 7, the bulk of umpulse events appear in the low frequency range, similar to the 6205 case. It should be noted that these events do happen in the 6206 data, but they do not peak above the expected amplitude of the baseline data as found in the 6205 and 6207 cases.

The fatigue case is a prevalent tested defect case in condition monitoring literature. Different methods are proposed to characterize the bearing damage relative to how damage is formed in industry. In this application, the bearing is damaged in such a way to resemble the grinding away of the bearing surface to mimic the damage accumulation. The exponential increase in bearing defect size is meant to represent the expected increase in the bearing vibration that occurs as damage accumulates. From the inner race defect, the changes from the baseline case are more gradual, rather than exponential, even in assessing just the 1 mm case from the baseline case. Further discussion is described in the bearing damage progression case in Section 4.1.3.3.

4.3.2.2 Comparison of Bearing Contamination

Bearings were contaminated based on the test plan provided in Section 4.2.2.2. As a recap from the section, bearing contamination has been found in applications ranging in size from 3 μm to 150 μm [480]. The particle composition ranges across a wide degree of abrasive materials (Silica, alumina, glass, iron oxide, *etc.*). The tested contamination amounts for oil-based applications range from 0.015 g/L to 2000 g/L [483, 87]. However, with grease contamination to the differences in

169
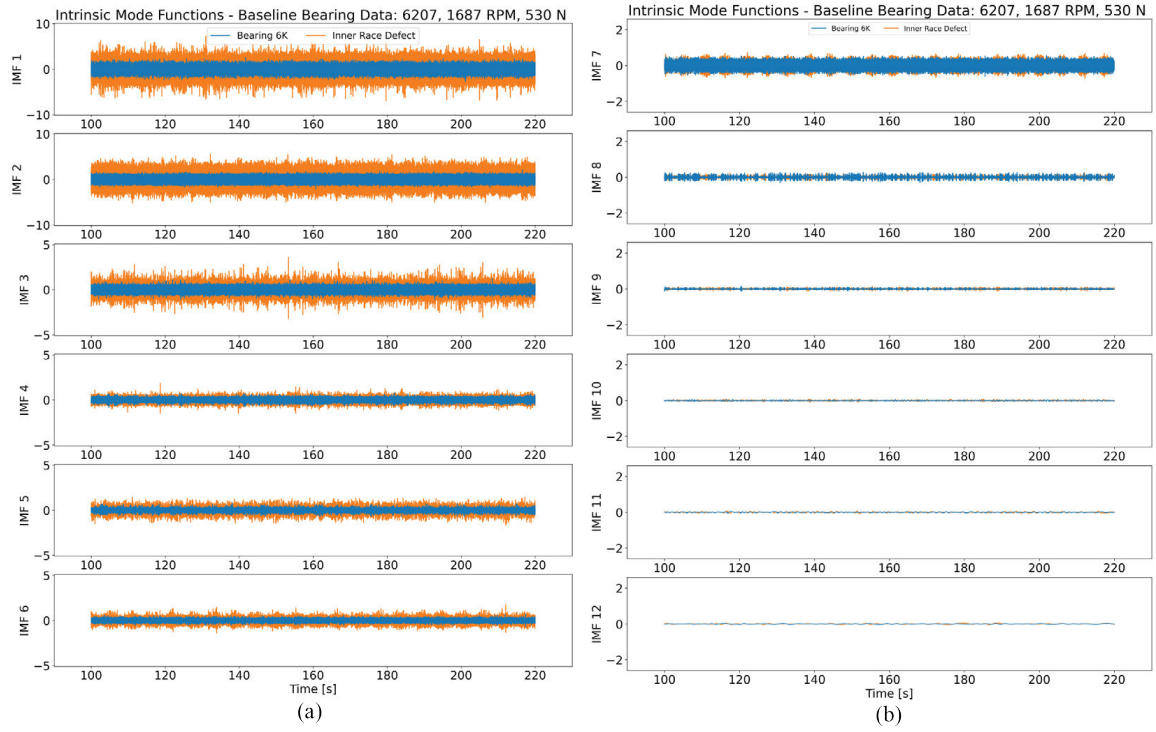
Figure 4.52: IMF representation of the 6207 bearings and 1687 RPM case for the steady state collections for the baseline (a) and inner race defect case (b)

viscosity and the intermixing of contaminates within grease, the tested concentrations are much higher, ranging from 5% of the grease weight to 45% of the grease weight [501, 502, 503].

Figure 4.53 shows the differences in the bearing surface after contamination has been applied to the bearing surface. In Figure 4.53 (a), the bearing surface still retains a shiny appearance, whereas in Figure 4.53 (b), the bearing ball surface appearance is duller. The difference in the ball appearance corresponds to smearing found along the bearing raceways. The smearing occurs due to the rolling elements sliding or skidding across contaminated material in the bearing raceway.

Figure 4.54 shows the change in the bearing raceway appearance as the bearing from the position in the load zone. Along each side of each picture is a scale for reference, where one dash is the equivalent of 2.54 mm. In the bearing outer raceways (Figure 4.54 (a), (c), and (e)), there is a smear that appears along the inner raceway of the bearing. The smear corresponds to the skidding and sliding of the bearing seen on the rolling elements. However, there appear to be flecks or spots that appear in the bearing raceway next to the bearing raceway. These deformations could correspond to the particles added to the bearing grease. The inner raceway for each bearing appears

170
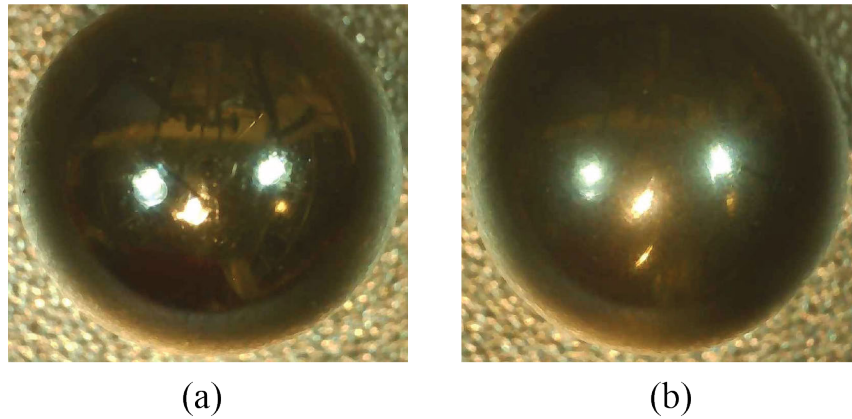
Figure 4.53: Change in bearing appearance after contamination added to grease. (a) Healthy bearing, (b) contaminated 6205 bearing from Contamination Code 20

to have suffered indentations from the particles added to the bearing raceway.

As the damage progressed to the next stage, the smearing on the outer raceway became more pronounced. Figure 4.55 shows the difference between smearing in the outer ring without magnification. In the 6205 bearing, the contamination appears multiple different spots along the bearing ring as the smearing occurs. In the 6206 bearing, the smearing appears along a larger portion of the bearing surface. The bearing surface smear appears more continuous and could represent a larger amount of the contamination material caught by the rolling element as it rotates around the bearing ring. In the bearings for contamination Code 24, smearing was found in all parts of the bearing raceway; however, it was more pronounced in the load zone.

Figure 4.56 provides an example of increasing damage for the 6205 bearing for bearing contamination. The inner race bearing damage in Figure 4.56 (a) does not appear to change significantly relative to Figure 4.54 in terms of pitting. However, scratches appear on the bearing surface elements in solid white circle. These scratches are reflected on the rolling bearing elements in Figure 4.56 (b) and (c). These scratches correspond to a similar axis relative to the ball spin axis. For Figure 4.56 (c), the pitting is shown to mix in with the scratches seen in Figure 4.56 (b). Figure 4.56 shows the outer race damage with the smearing in the dashed circles and the additional pitting and scratches in the solid white circle. the increase in pits and indents from Figure 4.56 (d) from Figure 4.54 (a) corresponds to the increase in the contamination amount. Hence, a larger amount of indentations were caused in the same duration. Additionally, in Figure 4.56 (d), there appear to be differences in the coloring along the bearing outer race smearing. These discolorations correspond

171

(a)

(b)

(c)

(d)

(e)

(f)
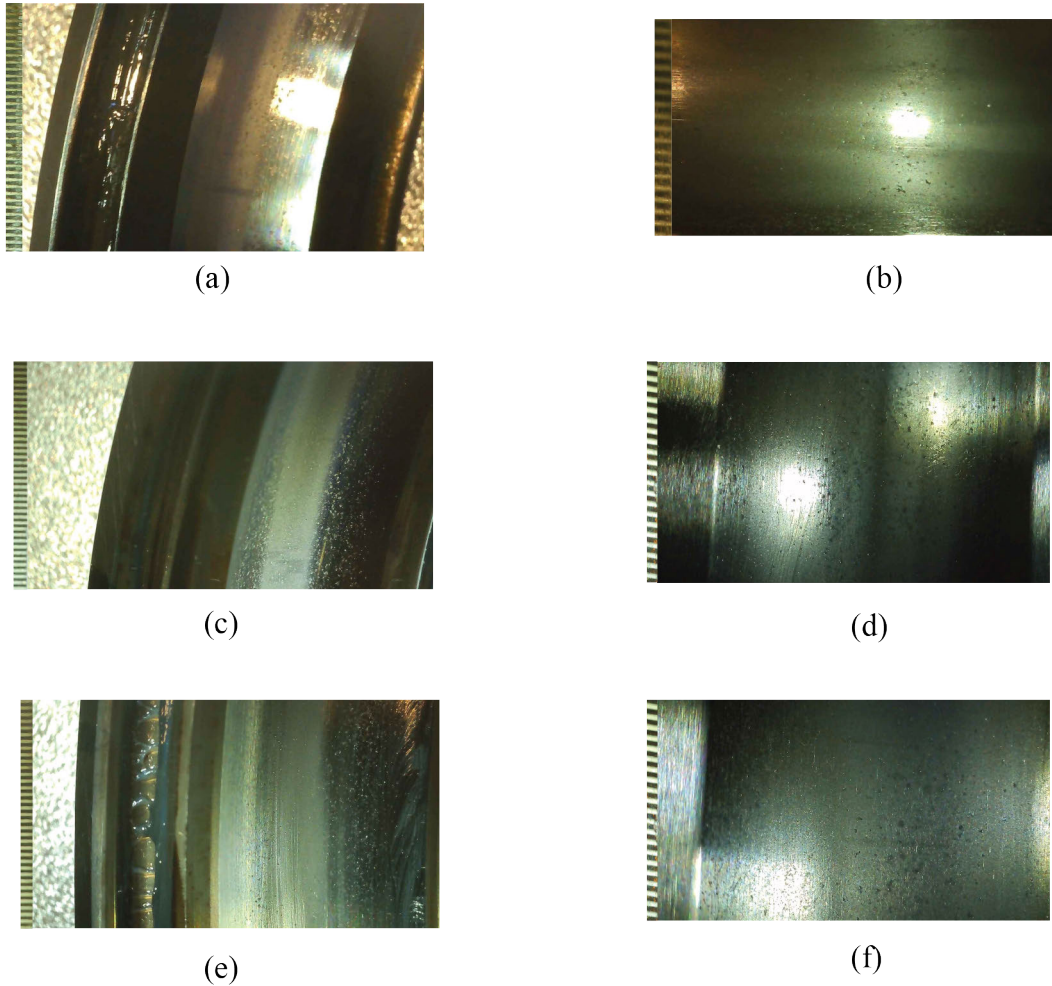
Figure 4.54: Change in contamination from different bearings. (a) Bearing 6205 Outer Raceway, (b) Bearing 6205 Inner Raceway, (c) Bearing 6206 Outer Raceway, (d) Bearing 6206 Inner Raceway, (e) Bearing 6207 Outer Raceway, (f) Bearing 6207 Inner Raceway

to the bearing contamination smearing. Figure 4.56 (e) shows a zoomed-in representation of the smearing. It is centered in the bearing ring and happens due to the bearing rolling element skidding and sliding along the bearing raceway.

The accumulation of wear on these bearings is considered at an accelerated rate. In real applications, the wear rate would gradually occur as more contamination particles are ingested. This testing aimed to compare the potential accelerated damage with the response and correlation to an actual bearing defect in the inner race case. Based on the bearing appearance, chemical and mechanical changes are present from adding contamination particulates. The tiny craters and smears that accumulate over time along the bearing raceway could replicate the effect of an accelerated
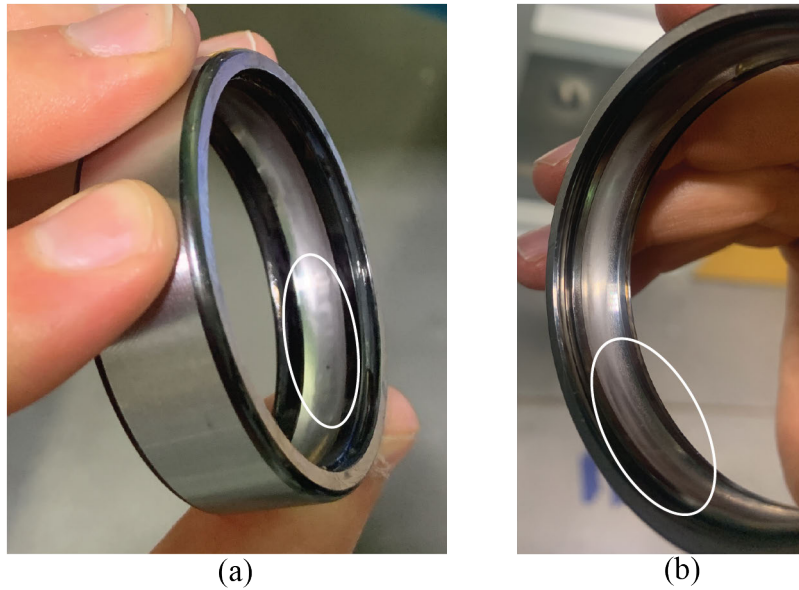
Figure 4.55: Contamination Code 24 smearing damage for 6205 bearing (a) and 6206 bearing (b)

failure of one crack over time.

Figure 4.57 shows the change in the bearing damage signature due to the increased contamination for the RMS and variance features. Similar to the bearing inner race defect case, the bearing damage propagates along a curve, where the bearing RMS increases exponentially in relation to the variance. The 6205 and 6207 bearing have similar changes rates for each case speed case. However, the 6206 bearing varies almost 3 times more exponentially in relation to the other two defect cases. Outside of the possibility of noise that could have affected the bearing results as it did with the inner race defect, it is possible that the bearing grease mixtures are not consistent with each other between each bearing test. The particles may "clump" and develop local concentrations due to the rolling elements.

Figure 4.58 represents the change in the bearing kurtosis and skewness for each different bearing size tested. The difference in the data shape between the baseline and contamination defect data. In this instance, the contamination data appears to focus exclusively on a skewness of 0. The centering of the bearing signature could stem from the impulse events relative to the baseline event. These impulse events occur based on the bearing rolling elements overrolling the bearing particles with the bearing raceway surfaces. The lack of variation in the bearing kurtosis could imply that the shape of the data does not change relative to the addition of the bearing particulate. The trend does
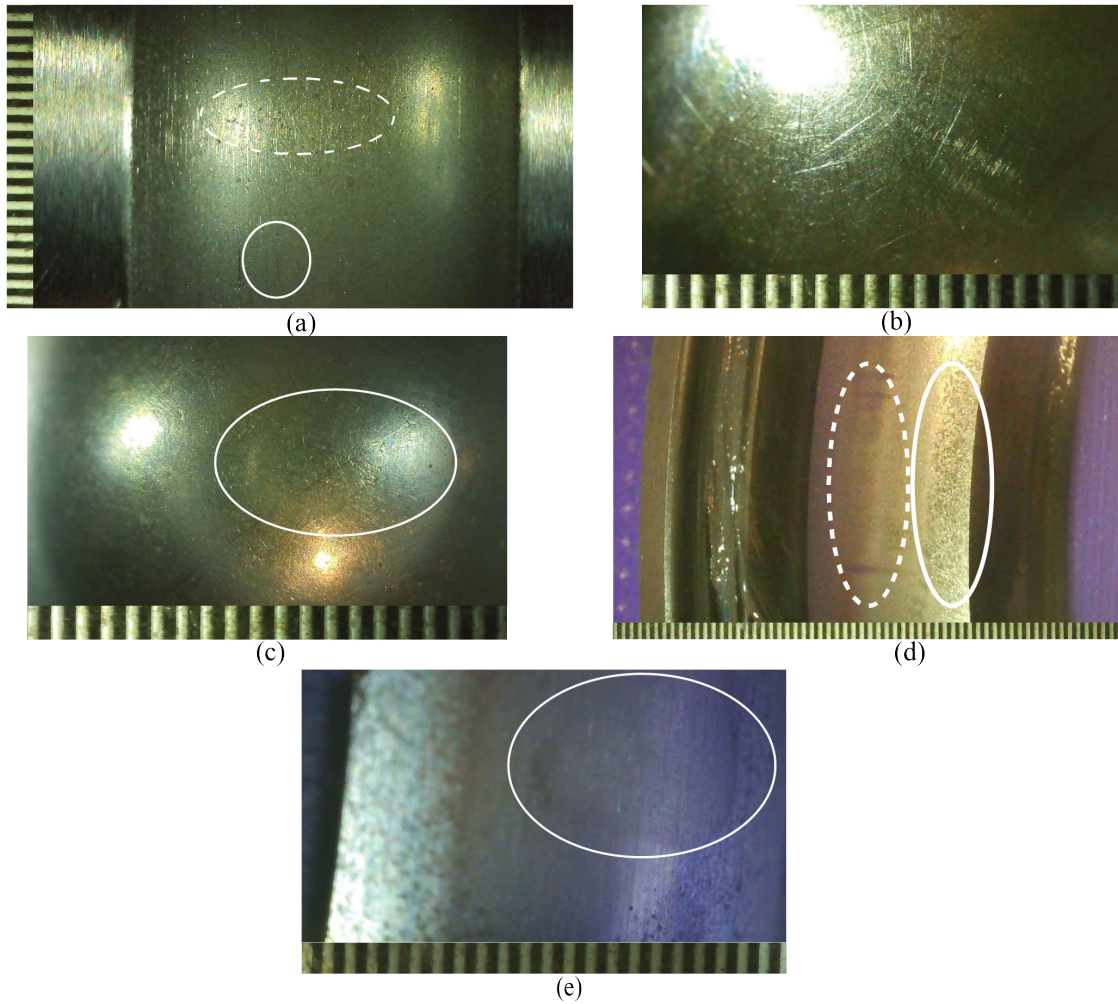
Figure 4.56: Contamination Code 24 for the 6205 bearing inner race damage with pits in the dashed circle and scratches in the solid circle (a), rolling element scratches in (b) and (c), outer race wear with contamination indentation along the side with smearing in the bearing grooves, and close up examination of smearing discoloration on outer race in (e)

not follow the expected occurrence as bearing damage increases, as found with inner race defects. The difference could stem from having a definite defect indent, whereas the bearing contamination particles do not allow for a complete defect shape.

From the initial defect state, the bearing damage differs from the inner race defect case between the contamination case based on observation. In the first instance, the RMS and the variance does not change as much in the inner race defect as it does with the contamination code case. The reason for the increase could stem from the overall elevated vibration amplitudes. The elevated amplitude could indicate multiple transient events recorded as the bearing particulates
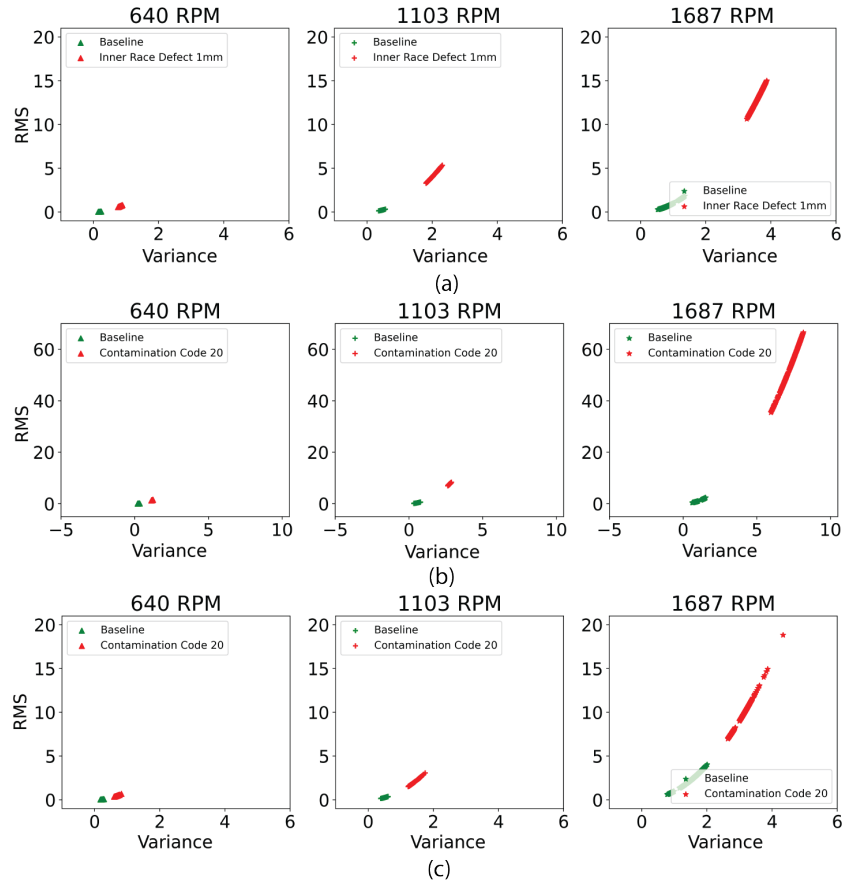
Figure 4.57: Bearing Code 20 RMS and Variance. (a) 6205, (b) 6206, and (c) 6207 bearing.

strike the bearing part surfaces. The change in bearing kurtosis points to a defined impulse event relative to the average bearing response at the 4th order, as long as the bearing standard deviation decreases or remains constant as the bearing amplitude increases. The bearing amplitude increases at all frequencies relative to the bearing impulse events for the contamination case and not at a particular instance.

Figure 4.59 considers the bearing FFT response for the 640 RPM case across the three different bearing sizes. For each bearing size, there is a general increase in the low frequency data, but the increase varies between the bearing sizes. In addition, there is a consistent increase in the high frequency vibration data across the bearing sizes. Compared to the inner race defect case shown in Figure 4.44, there is a greater increase in the high frequency range of the bearings for each bearing size. The contamination code has approximately three times the FFT amplitude over at least twice the range in each case. The range and height of the vibration amplitude support an overall increase
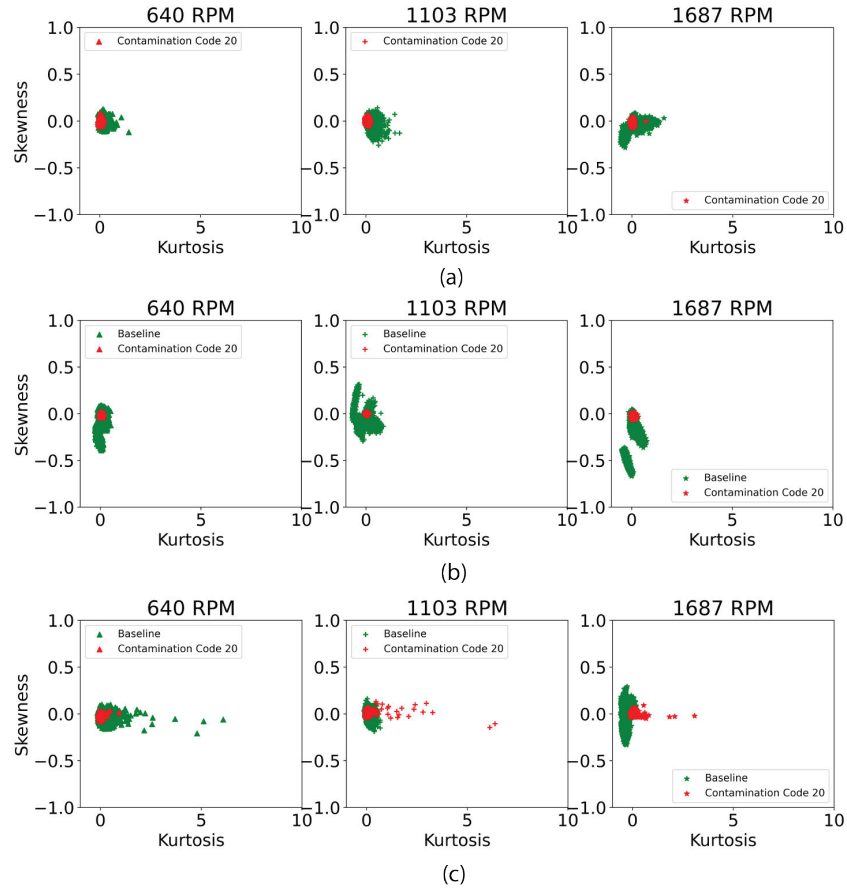
Figure 4.58: Bearing Code 20 Kurtosis and Skewness. (a) 6205, (b) 6206, and (c) 6207 bearing.

in the bearing vibration amplitude due to the contamination added to the bearing.

Figure 4.60 provides the change in the bearing FFT as the 1687 RPM case. There are increases in the lower frequency cases in the 6205 and 6206 cases, but not necessarily in the 6207 case. A similar occurrence was found in the inner race defect and could point to an anomaly in the baseline data for that case. There is an increase in the bearing high frequency data for each case that supports the general increase in vibration amplitude seen in the time series data. Similar to the 640 case, the change in vibration amplitude appears to happen over a wider frequency range than in the bearing fatigue defect case, supporting the possibility of multiple impact events on the bearing.

Figure 4.61 compares the bearing baseline data with a comparison of the contamination data in the frequency space for both the steady state and speed-varying cases. In this instance, the bearing data representation is normalized relative to the baseline max amplitude case. From
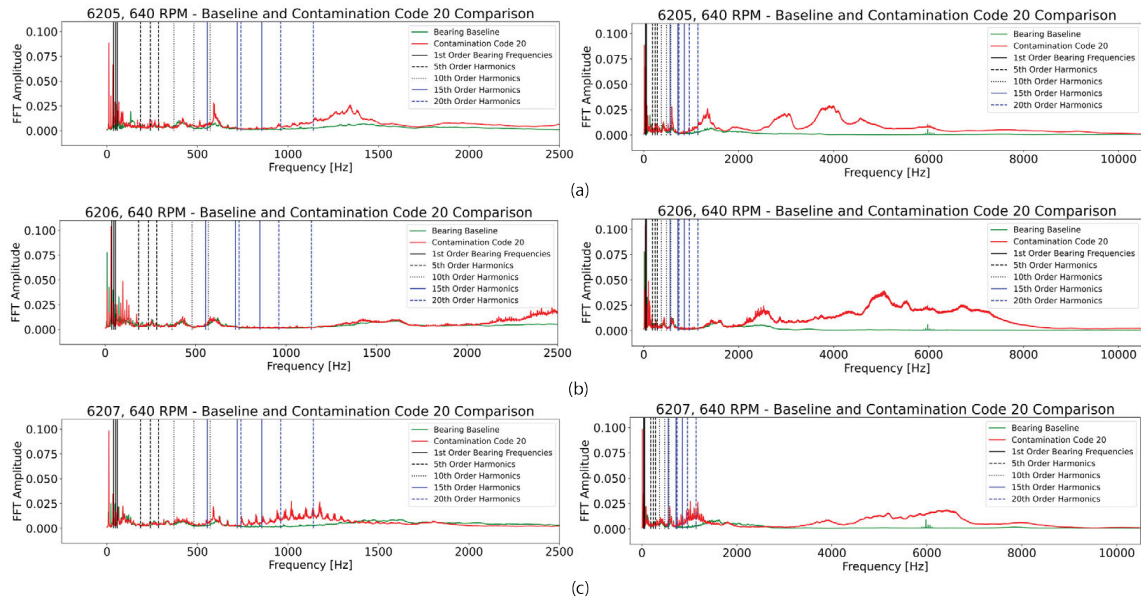
Figure 4.59: FFT at Bearing Speed 640 RPM for the 6205, 6206, and 6207 cases. Left column shows the frequency range between 0 and 2500 Hz; Right column shows the frequency range between 0 and 1000 Hz.
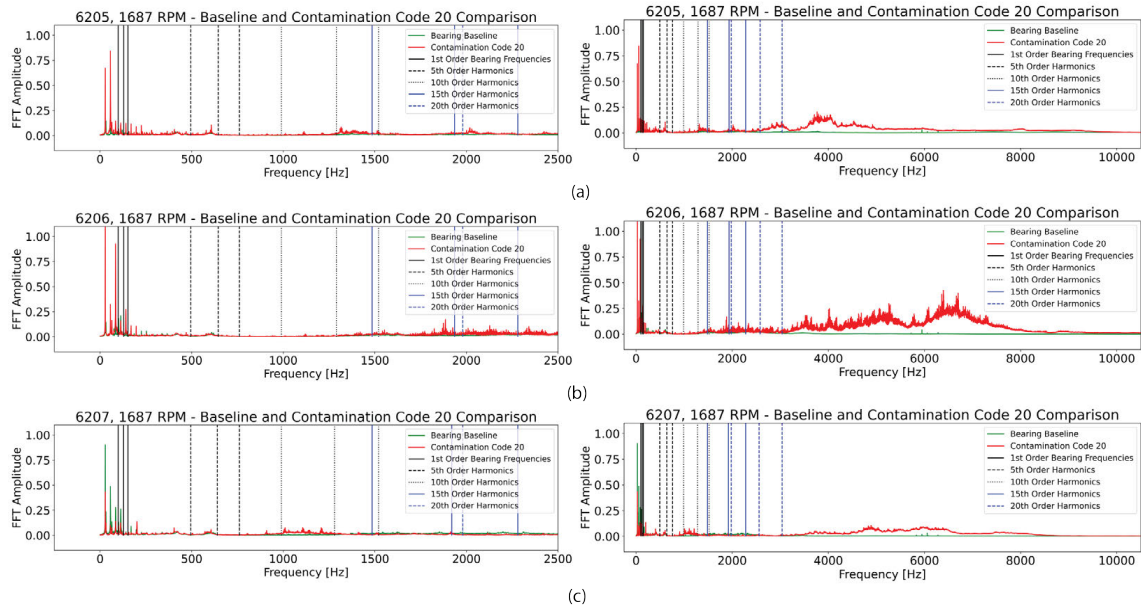
Figure 4.60: FFT at Bearing Speed 1687 RPM for the 6205, 6206, and 6207 cases. Left column shows the frequency range between 0 and 2500 Hz; Right column shows the frequency range between 0 and 1000 Hz.

the representation, though, it is clear that the bearing contamination case is much greater than either the steady-state or speed-varying cases. Change is apparent in low- and high-frequency cases;

however, that change is more prevalent in higher-frequency cases due to the darker discoloration. The case differs from the fatigue case, where there was a more moderate of high-frequency change compared to the contamination damage. In the speed-varying case, the effect on speed is more apparent in the low-frequency cases than in the high-frequency case for the bearing contamination case.
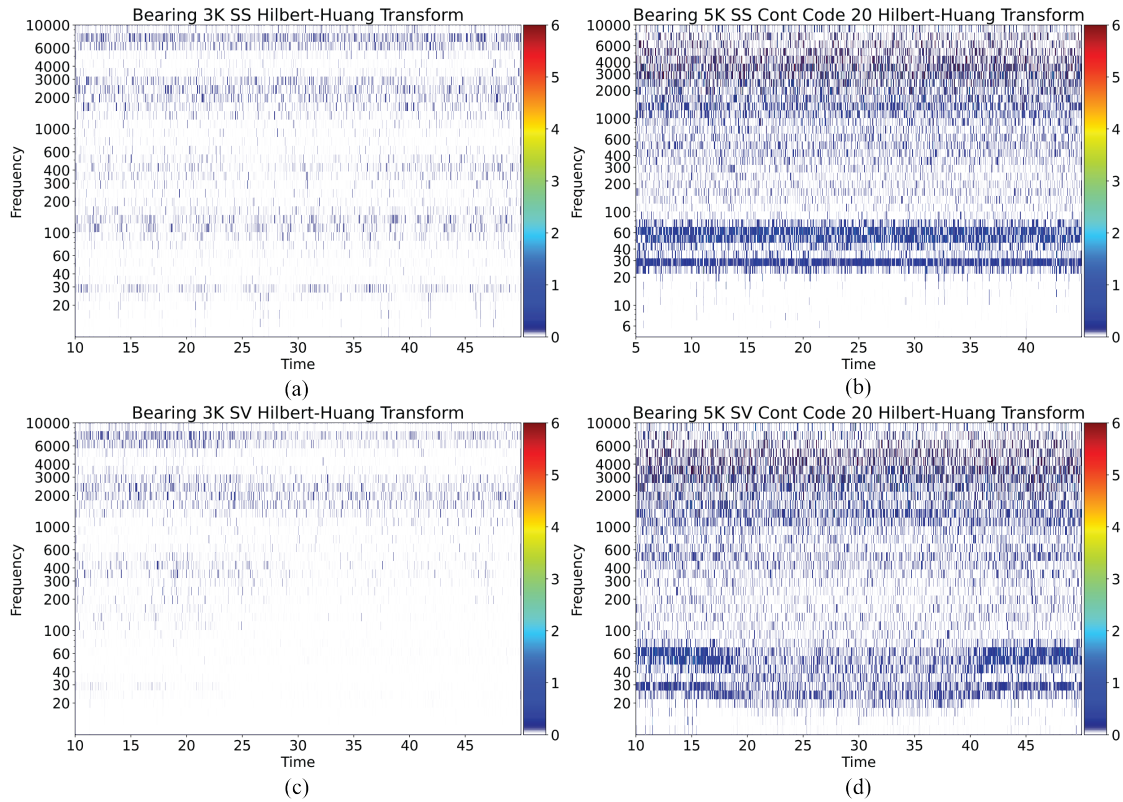
Figure 4.61: 6205 Bearing Response under speed varying and steady-state cases. (a) Steady-state baseline case, (b) steady-state contamination Code 20 case, (c) speed-varying baseline case, and (d) speed-varying contamination Code 20 case

Figure 4.62 shows a similar trend to the 6205 case. However, the amplitude is greater in the 6206 case than in the 6205 case. The increase in the frequency response matches the trend found earlier in the FFT. The primary difference appears to occur around the speed frequency at 30 Hz. The increased noise element at this range could stem from a resonance in the test system or from a misalignment in the coupler. Another possibility is the clumping of contamination particles during the data collection. The grouping of particles could mimic a similar event to a fatigue defect, except at the bearing speed frequency. The high amplitude in the speed-varying case does not

appear to drop relative to the amplitude normalization threshold relative to the baseline class. The accumulation in the low-frequency data explains the increase seen in both the RMS and variance from the time analysis. Between the different speed cases in the damage stage, the drop in amplitude could mimic similar trends seen in the baseline case. However, with the speed-varying bearing data normalized to the max case of the 1687 RPM data, there is still a clear separation in the high frequency data.
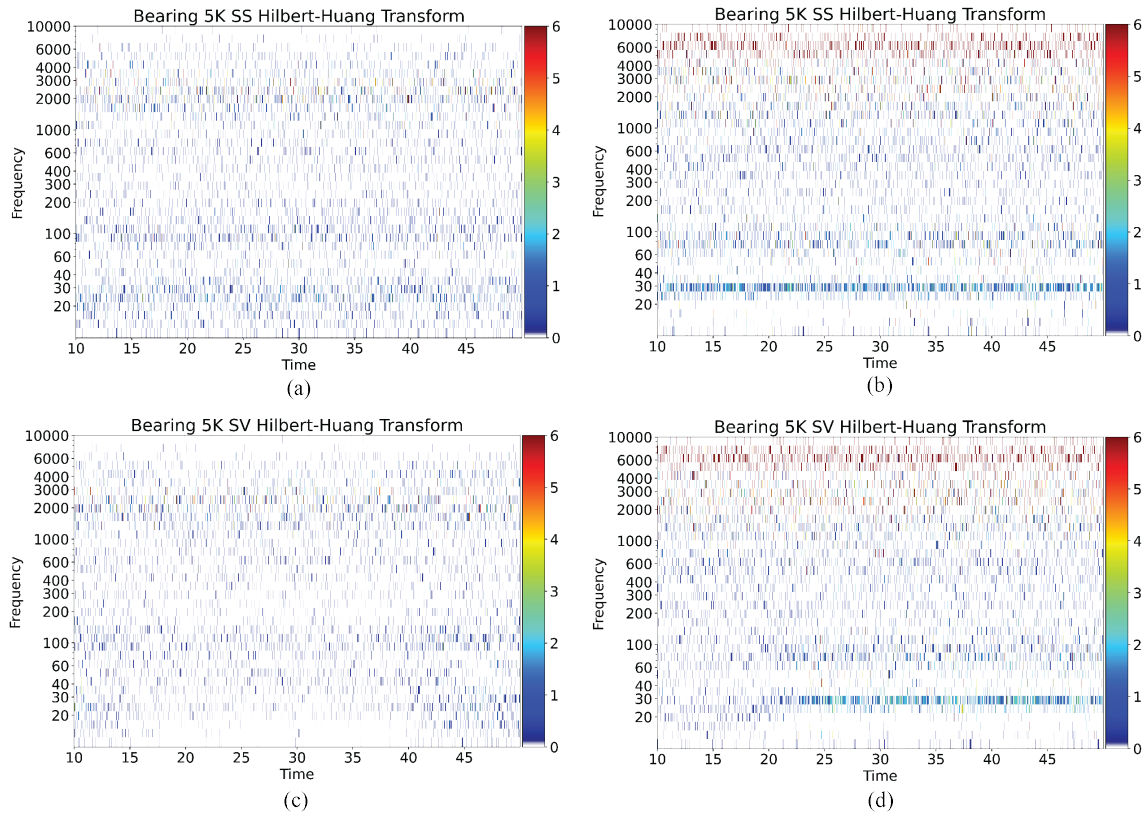
Figure 4.62: 6206 Bearing Response under speed varying and steady-state cases. (a) Steady-state baseline case, (b) steady-state contamination Code 20 case, (c) speed-varying baseline case, and (d) speed-varying contamination Code 20 case

Figure 4.63 shows the 6207 bearing damage case for both the steady-state and speed-varying cases in the Code 20 contamination case. The 6207 case follows a similar trend as expected from the 6205 case. However, there is a lower amplitude in the low-frequency data. A lower amplitude could stem from a more even distribution of particles within the grease or the lower possibility of noise and resonance related to the speed frequencies. Incidentally, the higher frequency data maintains a large amplitude consistent with the other bearing cases. In the speed varying case, the speed change

179

does not appear to affect the high amplitude readings relative to the baseline case.
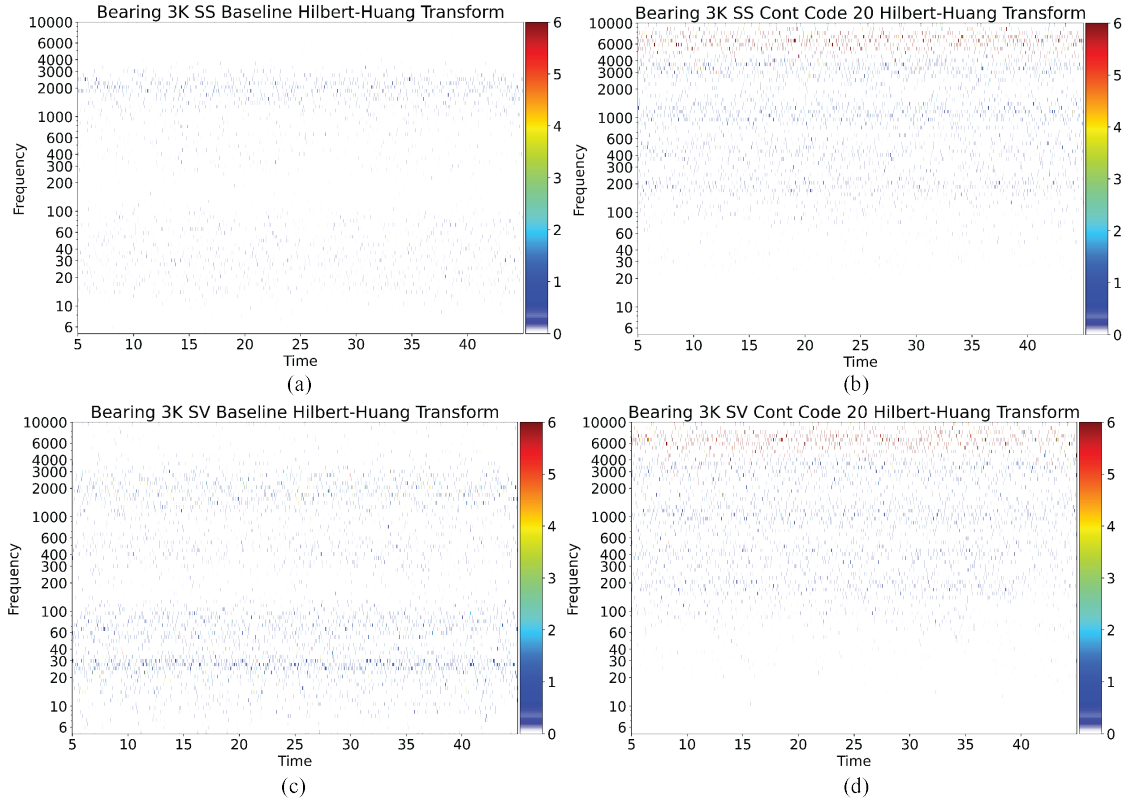
Figure 4.63: 6207 Bearing Response under speed varying and steady-state cases. (a) Steady-state baseline case, (b) steady-state contamination Code 20 case, (c) speed-varying baseline case, and (d) speed-varying contamination Code 20 case

Figure 4.64 shows the IMF readings comparison between the baseline and contamination Code 20 vibration signatures for the first 12 IMFs. IMF 1 through 5 has a much higher amplitude relative baseline case. This trend is supported based on the damage modes seen with the inner race fatigue defect. However, the bearing fatigue amplitude case remains greater than in the low-frequency data compared to the baseline data case. Incidentally, the primary difference between the inner race fatigue and the contamination case is the lack of a defined defect area, such as an engraving, dent, or knock on the bearing raceway. Additionally, this is reflected in the low-frequency IMFs by not having the expected impulse events related to the bearing frequencies. For the speed-varying cases, the speed changes see a more drastic amplitude change in the high-frequency data versus the low-frequency data. The speed-varying speed changes appear captured in the IMF 1, 4, and 5 cases for the contamination case, whereas in the baseline case, the speed changes are captured

in IMF 1, 4, 5, 6, and 7. Due to the changes occurring in different IMF signatures, selecting and focusing on specific IMF bands related to these frequencies can separate healthy versus damage signatures.

In the contamination case, damage is inflicted based on environmental conditions versus purposeful damage. The characterization of the bearing surface shows how the bearing damage is accumulated across the entire bearing surface versus a specific section as is shown with the fatigue case. Consequently, the contamination case shows a much higher amplitude reading as opposed to what occurs with a defect case on the inner race. The impact of the bearing rolling element with each of these individual particles causes further asperity to the bearing surface at a much quicker rate.

4.3.2.3 Comparison of Bearing Lubrication

In the lubrication case, the aim was to create bearing damage by increasing the amount of friction between the bearing rolling elements. The increased friction causes a buildup of heat and stress between the bearing rolling elements forming early wear on the bearing surface. Incidentally, there are instances where too much grease causes burning and an increased heat buildup during the bearing run time. The increased heat buildup from too much grease can cause early bearing degradation by impeding the bearing rolling element path. During the experimental stage, it was found that the grease in the bearing in the initial case was less than the expected grease amount. In this instance, the initial lubrication case is an over-lubrication case, whereas the middle lubrication case represents a similar lubrication case to the expected bearing lubrication.

Figure 4.65 contains the lubrication features concerning the RMS and variance comparison and Kurtosis and Skewness comparison across the three different bearing sizes. Due to issues in the data collection, only these data classes are available for each bearing. In the 6205 case, there was no noticeable and measurable difference between the different lubrication cases in the RMS and variance data. However in the, kurtosis and skewness data, there is more variation that occurs. The change in kurtosis and skewness could signify a shift in the data distribution to the changing data states.

For the 6206 case, the first lubrication at 25% comes through in the 640 and 1103 RPM cases. In this instance, there is a sudden jump in the RMS and variance data relative to the initial lubrication cases at 75% and 50%. However, despite the spike in the data, there was little change in
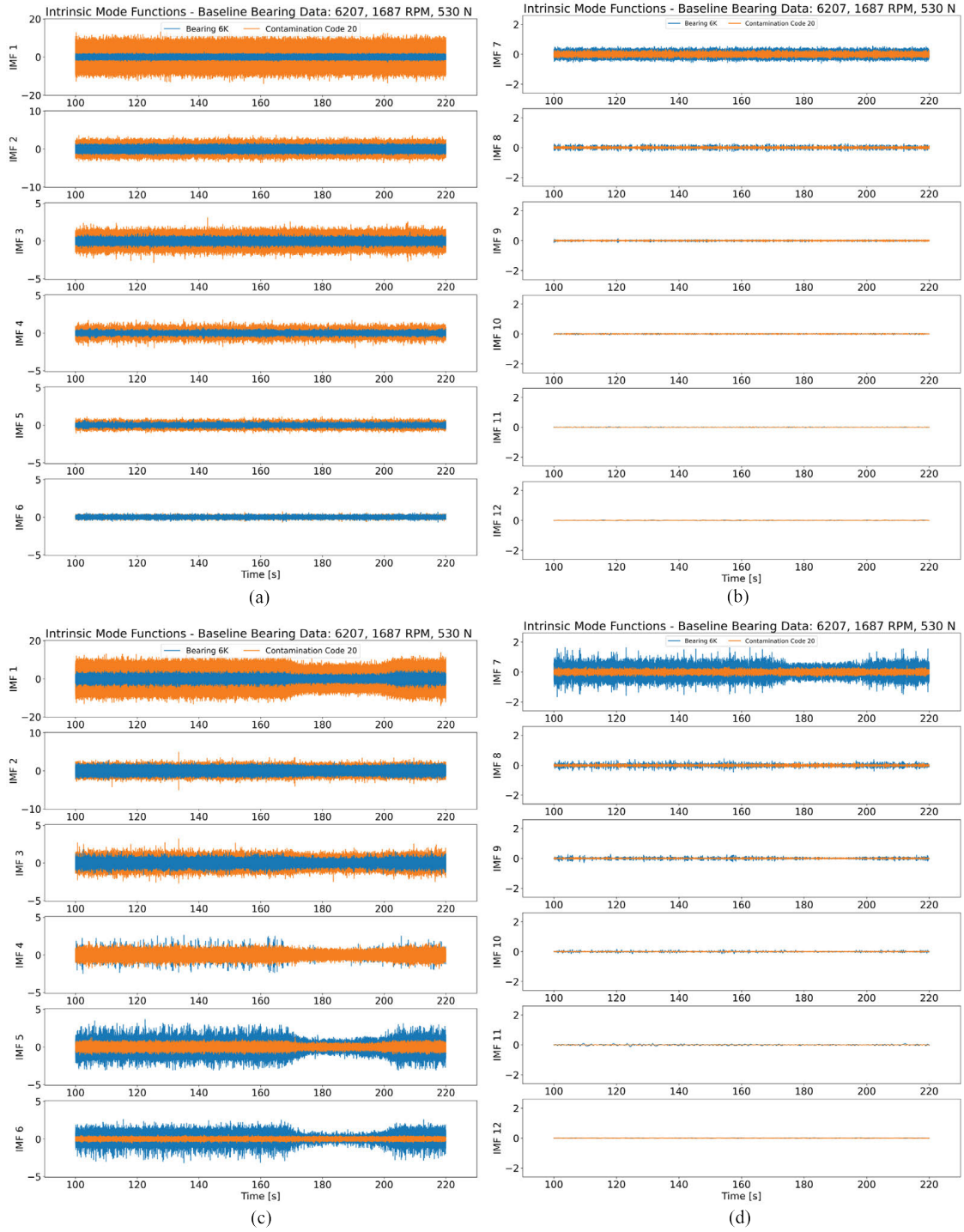
Figure 4.64: 6207 Bearing response from the IMFs under speed-varying steady-state cases. (a) Steady-state baseline case, (b) steady-state contamination Code 20 case, (c) speed-varying baseline case, and (d) speed-varying contamination Code 20 case
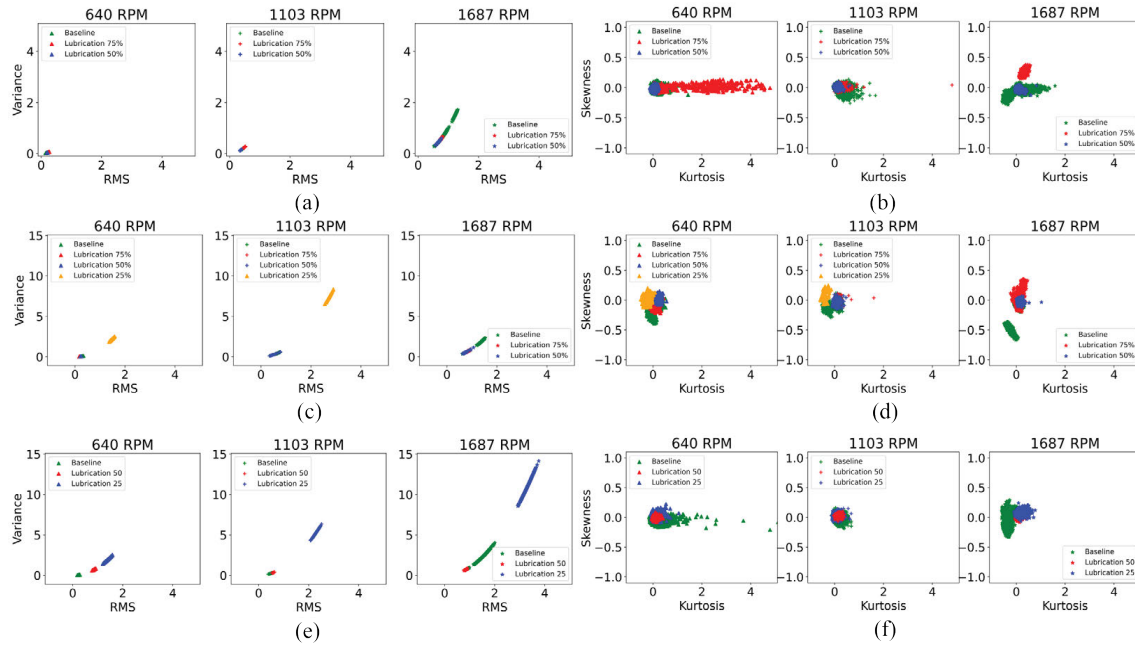
Figure 4.65: Changing bearing lubrication across the different time features cases and bearing sizes. (a) 6205 RMS and Variance comparison, (b) 6205 Kurtosis and Skewness comparison, (c) 6206 RMS and Variance comparison, (d) 6206 Kurtosis and Skewness comparison, (e) 6207 RMS and Variance comparison, and (f) 6207 Kurtosis and Skewness comparison

the kurtosis and skewness. Based on this observation, this could mean that the general amplitude for the bearing data increased, but the distribution in the data did not change. In the 6206 case, the lubrication case appears closer to the bearing baseline case based on the data shape.

For the 6207 case, all three speeds have the lower lubrication case collected. In this instance, the lubrication 25% case greatly outweighs the baseline case. In the 640 RPM case, the 50% lubrication case appears to separate from the baseline case. In this instance, it is possible that the bearing had more difficulty passing through the grease at a slower speed than in the higher-speed cases. However, this trend did not align with the other bearing cases. The higher response could be due to the larger rolling elements relative to the other bearing sizes. The Kurtosis and skewness between the different lubrication states are similar to the other bearings, minus the 6205 640 RPM case. This result could be due to potential outliers during the data collection procedure that should have been considered.

Figure 4.66 contains the frequency responses related to the different bearings tested at the different speed cases. From the 6205 graphs, the bearing signatures remain at a similar variation as seen with the baseline conditions. In the 6205 640 RPM case, there are some slight variations

around the 500 Hz frequency. However, these frequencies do not appear to transfer to higher-speed cases. It is possible that at the higher speeds the rolling elements can pass through the grease easier, which was observed in the opposite view for the 6207 bearing in the time series data.
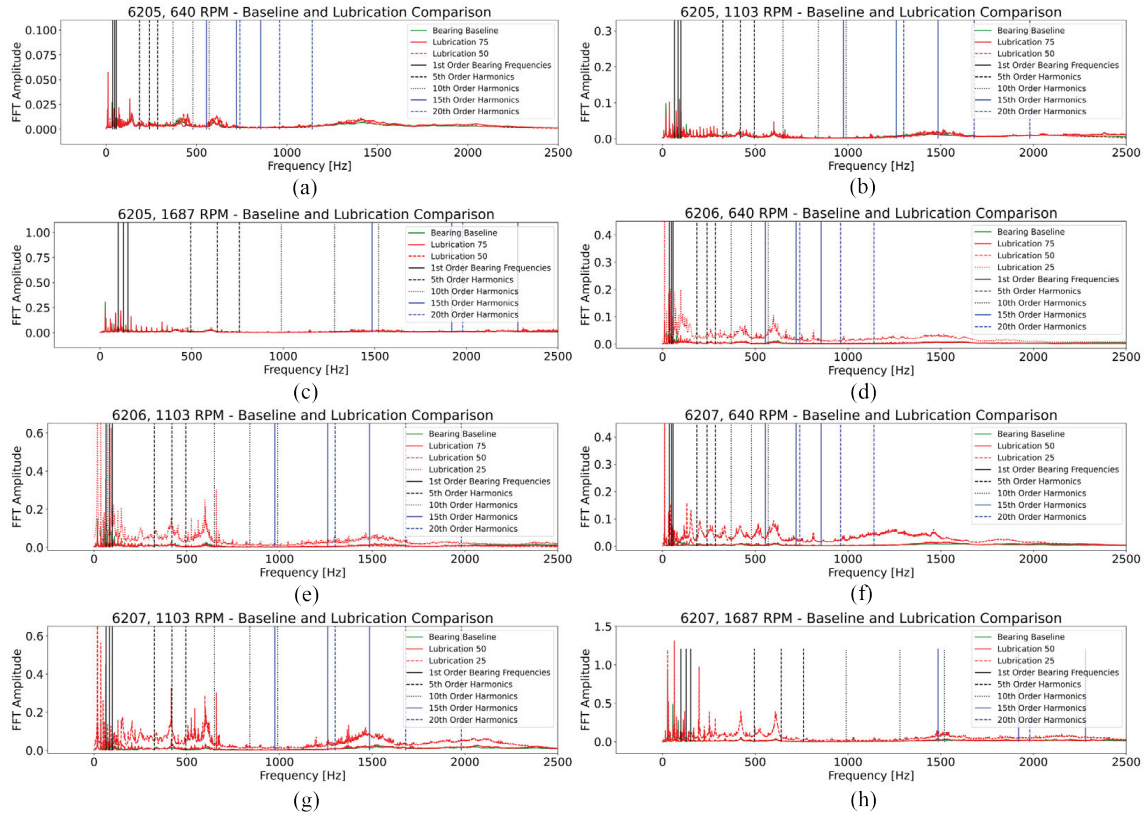
Figure 4.66: Changing bearing lubrication across the different bearing size and speed cases. (a) 6205 640 RPM case, (b) 6205 1103 RPM case, (c) 6205 1687 RPM case, (d) 6206 640 RPM case, (e) 6206 1103 RPM case, (f) 6207 640 RPM case, (g) 6207 1103 RPM case, and (h) 6207 1687 RPM case

In the 6206 case, the lubrication 25 data appears to jump up compared to the lubrication 75% and 50% cases, respectively. The jump could correspond to the increase in the bearing vibration amplitude relative to the different defect stages. A similar increase is found in the 6207 bearing data. The overall increase in the FFT response appears to correspond to an increase in the bearing vibration amplitude. The spikes in the data appear to correspond to the bearing frequencies and their subsequent harmonics. The registered spikes could correspond to an increase in the bearing friction as the rolling elements roll over different points of the bearing raceway. The increase in bearing friction could then correspond to an increase in damage. The damage would eventually form on the bearing raceways and rolling elements as fatigue damage.

Figure 4.67 contains the IMFs for the 6207 IMF case for the steady-state and speed-varying conditions. For the steady-state condition using the EMD method, the high frequency IMFs actually have more of the change in data content, instead of the FFT, as was seen earlier in Figure 4.66. The primary difference could stem from the overall increase in amplitude for the 25% lubrication stage, relative to the baseline conditions. By IMF 5, the 25% lubrication stage shares similar responses as to what was expected by the baseline case. Interestingly, the lubrication 50% stage also decrease over time, but is considered less than the baseline case at this speed.

In the speed-varying case, more modulation occurs relative to the change in speed during testing. Interestingly, there are more impulse events in the higher frequency spaces (IMFs 1-3). The impulse events appear to correspond more to changes in speed rather than changes in the bearing damage state. However, considering that these disruptions also occurred in the contamination event at higher frequencies, changes in the lubrication patterns may affect the bearing responses in the high-frequency stages. The possible change in the impulse events could further signify a difference between the contamination stages in the diagnostic stage.

4.3.2.4 Comparison of Bearing Brinelling

In the Brinelling case, the damage was induced by dropping the bearings repeatedly from predetermined heights. The method was meant to represent a scenario where the bearing is incorrectly installed onto the bearing piece of equipment. SKF [7] considered the probability of improper installation as 16.6% of all possible failures. However, due to their high precision nature, it is possible for the bearing to be installed under improper conditions. Improper installation could apply when the bearing is dropped, pressed on an improper shaft, or installed incorrectly. Morales-Espejel et al. [486] utilized an artificial indenter to mimic these improper conditions to create more similar fatigue cases. However, in this instance, the interest is focused on the possible signal changes before the damage propagates to the fatigue cases, where there is visible damage on the surface. Other methods that could have been used to simulate the damage is using a press to force an indentation of the ball on the bearing surface. This method was not used to avoid any unforeseen damage to the bearing surface

Figure 4.68 and 4.69 show the bearing damage cases relative to the different bearing sizes. In this instance, bearing damage can vary based on the size of the ball and the angle the bearing is dropped. Additionally, the bearing damage could be scored by any particles that enter the bearing

185
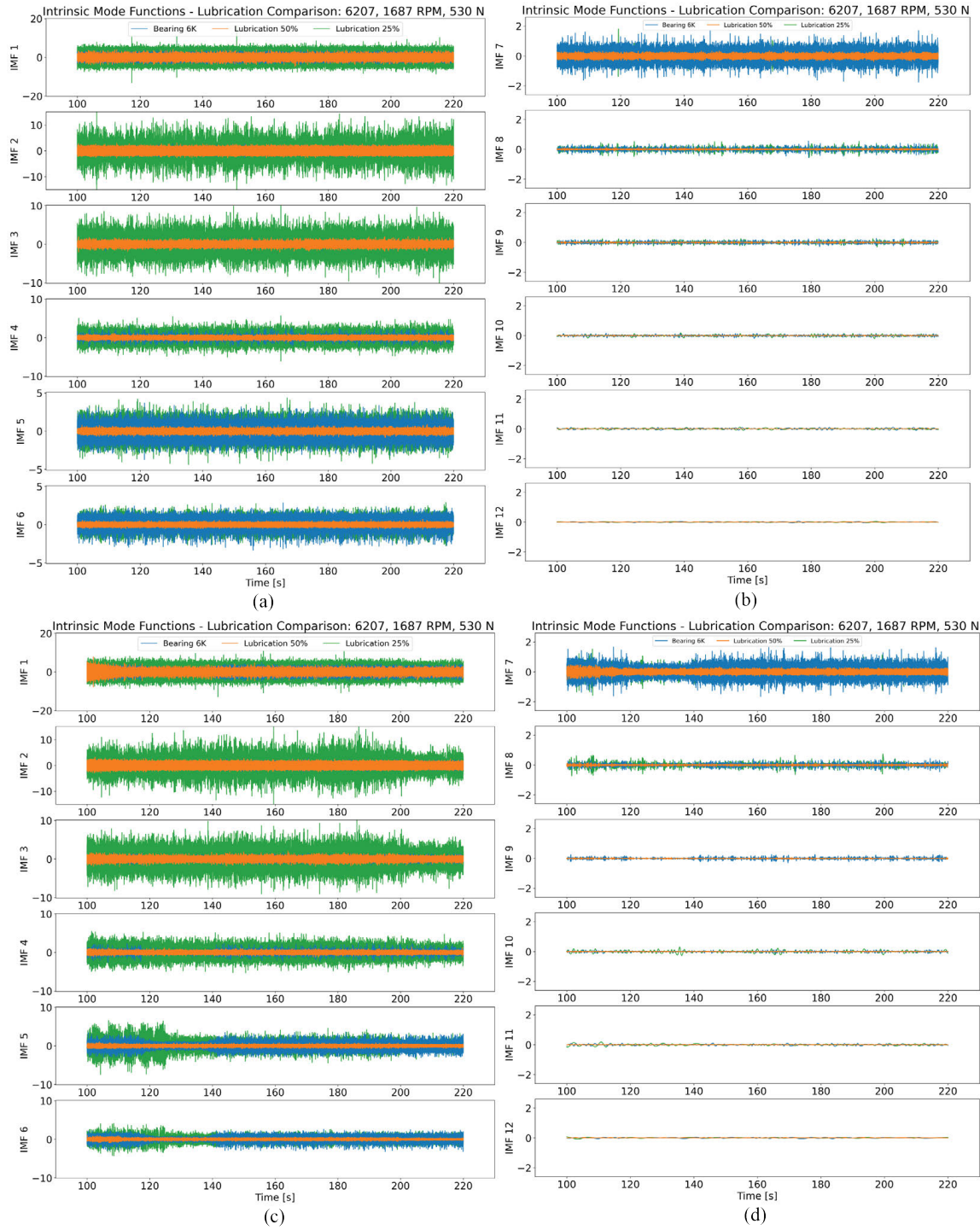
Figure 4.67: Bearing 6207 Lubrication case at 1687 RPM for the steady state (a-b) and speed varying cases (c-d)

raceway and strike the surface. Hence, the variation between the 6205 bearing and the 6206 damage. It is possible that during the bearing drops, the rolling element rotates around on its axis along with a foreign material to cause scratches on the bearing surface. Even in the 6206 case, faint scratch marks appear on the bearing surface. These are not as apparent on the bearing surface, meaning there could have been a lack of a foreign element or a slight difference in position to cause the lack of physical damage.

(a)   (b)
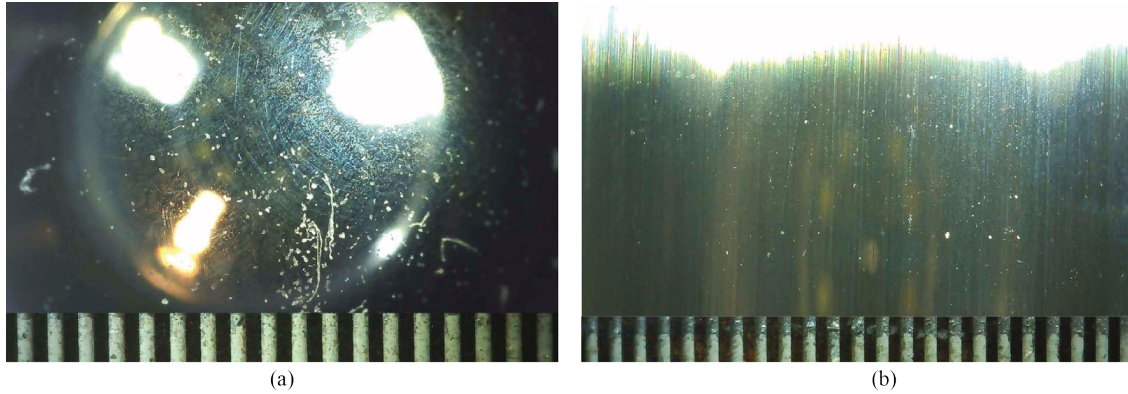
Figure 4.68: 6205 bearing Brinelling case after dropping the bearing at 1 m

(a)   (b)
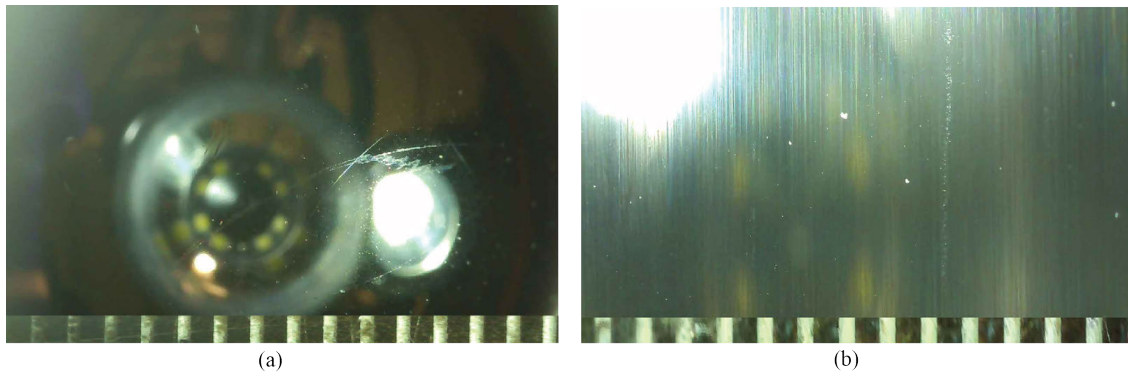
Figure 4.69: 6206 bearing Brinelling case after dropping the bearing at 1 m

Figure 4.70 highlights the difference in damage signals across the different bearing damage cases in the 0.5 m and 1.0 m Brinell cases. The strike from the Brinell surface between the two bearing raceways and the rolling elements should leave a slight indentation on the bearing surface and signal a difference in variation in the bearing surface. However, due to the amount of variability seen in the bearing data, the variability in damage is not distinguishable based on the time-series analysis for the 6206 and 6207 cases. There is some amount of separation in the bearing 6205 case;

however, the difference could be captured due to the additional scratch marks seen in on the bearing element versus the bearing indentation. Another anomaly that could characterize the difference is the progression of damage. For example, in the 6205 case, the 0.5 m has a higher amplitude than in the 6207 case, signifying a potential issue with the test stand configuration during testing. The frequency analysis is used to verify what the potential issue with the damage could entail.
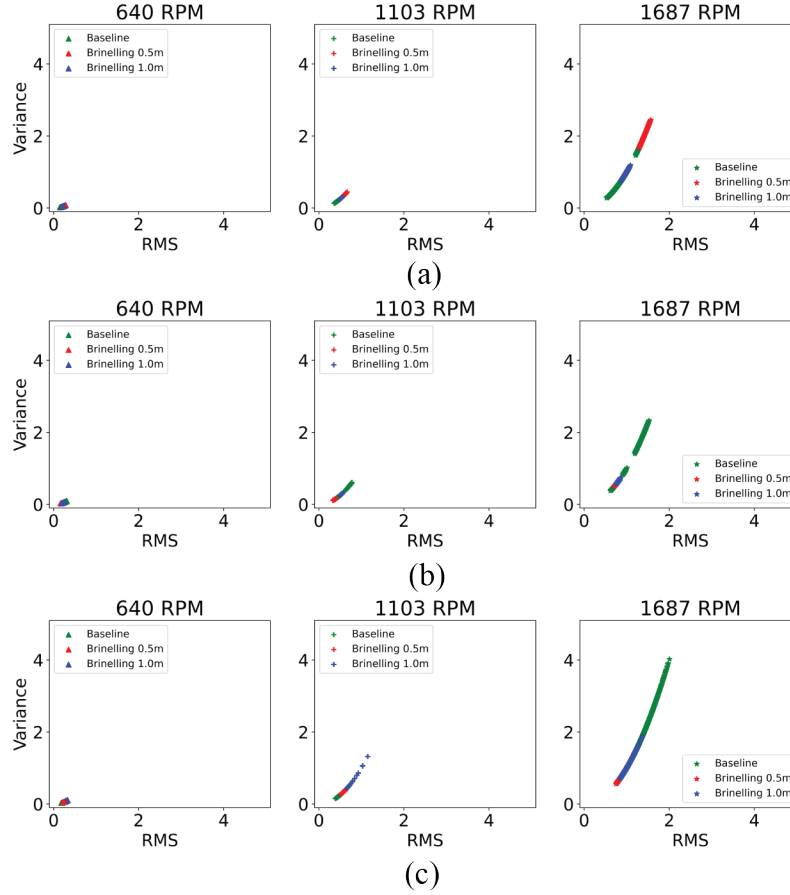
Figure 4.70: Bearing RMS and variance relative to the different bearing sizes. (a) 6205, (b) 6206, and (c) 6207 bearings

Figure 4.71 shows the representation of the kurtosis and skewness for the 6205 and 6206 bearings. This representation provides a different representation of data changes during the data generation. From the data generation, the 1.0 m stage is slightly greater than the 0.5 m stage. The difference in the damage case could also be tracked through the changes in skewness in either case. However, those changes may relate more to the overall system than the actual bearing. However, due to the intermixing of the bearing baseline data with either case, it is possible that there is not

an established damage trend possible with the collected data based on these initial readings. In other damage cases, the data shape changed as the damage was inflicted on the bearing. Hence, because of the lack of change in the data distribution shape and the low degree of change seen in the RMS and variance cases, there may not be any discernible damage applied to the bearing. However, the frequency analysis is still considered to verify if there were any relative changes in the bearing response.
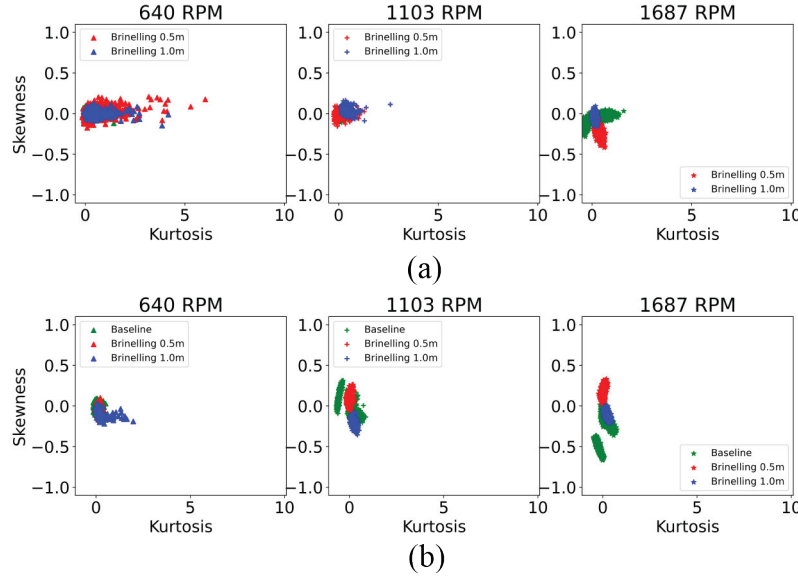
Figure 4.71: Bearing RMS and variance relative to the different bearing sizes. (a) 6205 and (b) 6206 bearings

Figure 4.72 considers the 6205 bearing speed case based on the FFT representation. From the FFT representation in the 640 RPM case, there is a larger spike at just above the bearing characteristic frequencies. For the other two speed cases, it appears that the frequency space representation for the Brinelling case exceeds the baseline case in the low-frequency range. This observation deviates from the other damage stages, where the initial bearing damage appeared to occur at higher frequencies versus the lower frequencies. There was not a substantial deviation for the high-frequency data, as was seen in the contamination and inner race defect cases. The lack of high-frequency signals could correspond to an impulse event related to a physical defect, as found in the inner race or contamination case.

Figure 4.73 compares the bearing Brinelling cases in the steady-state and speed-varying cases across their 12 IMFs. Contrary to the other damage cases, the low-frequency representations

189
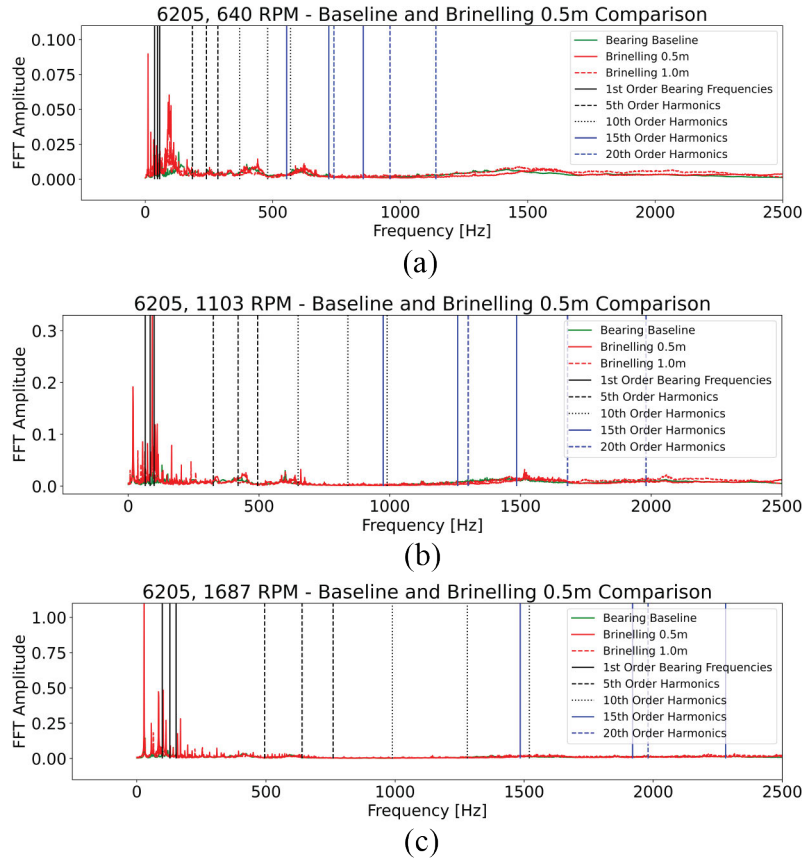
Figure 4.72: 6205 Brinell case for the 0.5 m and 1.0 m cases under different speed cases: (a) 640 RPM, (b) 1103 RPM, and (c) 1687 RPM

contain more anomalous or damage signatures than in the baseline case. The IMFs between 4 through 7 appear to correspond to the frequencies between 60 and 1000 Hz. Impulse events appear that could correspond to similar occurrences as seen in the inner raceway fatigue case. Earlier, it was thought that the lack of presence in the high-frequency data could not be present due to the lack of impulse events. However, there are low-frequency events based on the EMD analysis. Hence, the high-frequency data must be caused by the change in surface roughness and the number of strikes against the bearing surface as it occurs in the contamination and fatigue case. Incidentally, the increase in the bearing defect damage for the Brinelling case could correspond more to a smooth but recurrent indentation.
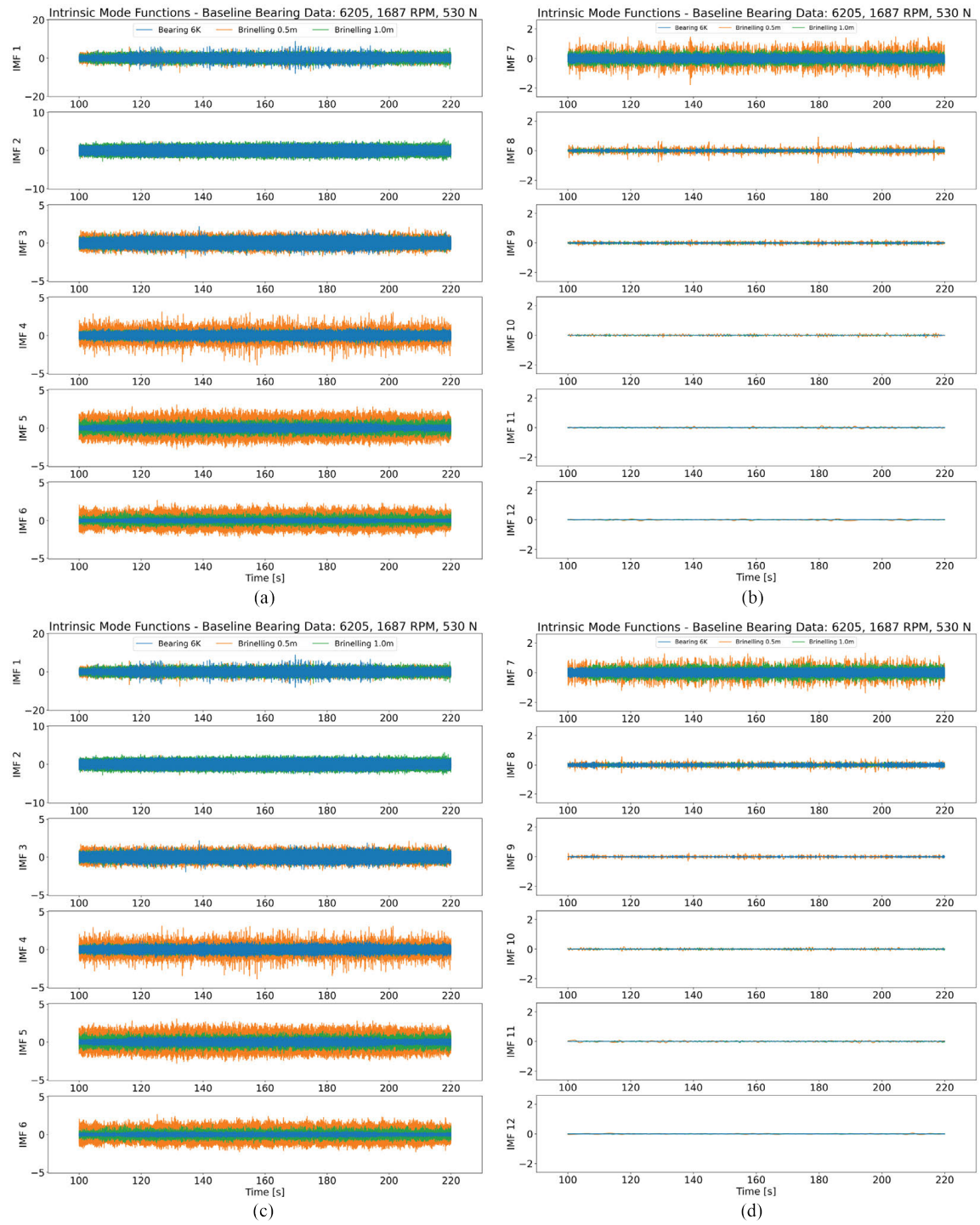
Figure 4.73: 6205 Brinelling damage case for the (a, b) steady-state cases and (c, d) speed-varying cases

4.3.2.5 Progressing Failure Conditions: Fatigue & Contamination

Each failure mode is collected over progressing stages of failure by increasing the amount of damage applied in the case of fatigue or contamination, the amount of dents from the brinelling, or the amount of lubrication removed from the purposeful lubrication. However, in the empirical observations, changes in damage were only significantly noted in the fatigue and contamination. The most likely cause for the progressing damage stems from the surface damage caused to the bearing raceway surface due to the actual deforming of the surface.

However, the surface damage does not necessarily mean a clear increase in the defect size as it grows. Take for example, Figure 4.74 and Figure 4.75 which show the increase in the average FFT relative to the different defect stages compared to the baseline state. In the 1103 RPM case (Figure 4.74, the 6205 bearing appears to follow a discernible trend, where as the damage grows the low frequencies increase in amplitude. However, in the 6206 case, there is a slower growth in the low frequency amplitudes. However, in the high frequency range, there are more discernible differences as the bearing defect changes. There also appears to be some anomalous spikes in the bearing data for the 1103 case, that could symbolize potential noise from the system.

In the 1687 RPM case (Figure 4.75), the bearing data appears to fluctuate between the different classes more so than in the 1103 RPM case. In the 6205 case, the bearing data for the 1mm and 3 mm case appear to be interchangeable with the baseline data at some frequencies. For this reason, it is possible to notice that the baseline data intermixes with bearing inner race defect, making it harder to differentiate between the two different methods. A similar occurrence happens in the 6206 and 6207 data as well, it is possible that the intermixed data could cause confusion in the classification of the different defect sizes and the progression.

A possible reason for the difference in the bearing configuration is the result of unaccounted for elements in the test stand such as looseness or running equipment in the vicinity. Hence it is possible that during the different bearing change-outs something was modified that caused for an incorrect reading during testing. The possible variation is further investigated during the physical modeling for the test system. Similar variation could also come from the different people conducting the testing in the system.

Figure 4.76 contains the average FFTs for every contamination stage for the different bearing sizes. In this instance, the increase in damage is apparent between the different stages. However,
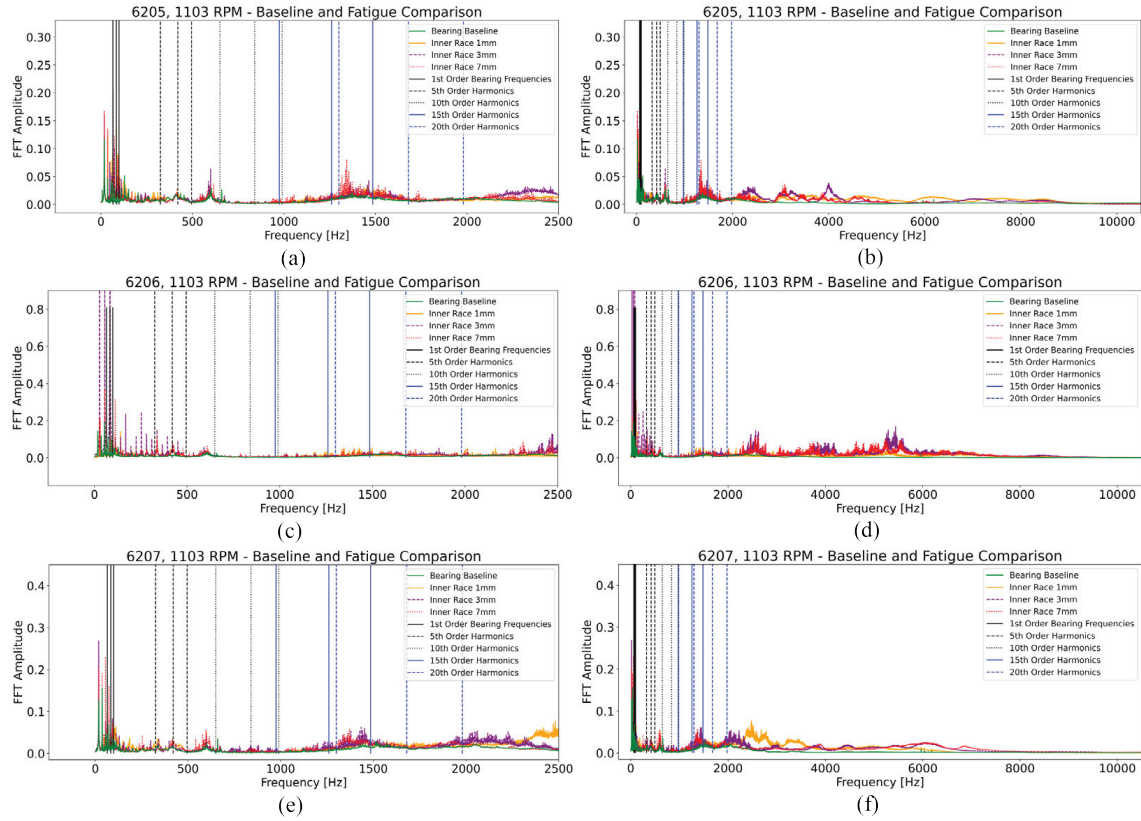
Figure 4.74: 1103 RPM case for the three different bearing sizes at the low frequency and full frequency ranges: (a) 6205 0 Hz - 2500 Hz, (b) 6205 0 Hz - 10000 Hz, (c) 6206 0 Hz - 2500 Hz, (d) 6206 0 Hz - 10000 Hz, (e) 6207 0 Hz - 2500 Hz, and (f) 6207 0 Hz - 10000 Hz

there are still noticeable shifts in the frequencies for the 1103 6205 case, where the initial contamination case has a higher amplitude in the higher frequency, but then begins to shift to the lower frequencies as more contamination accumulates. The lasting damage is then shown concerning the clean grease case tested without contamination ("Code 0"). The instances of variation appear to vary based on the different classes. In the case of the 6205 bearings, the lasting damage appeared as large as the heavy contamination case. However, in the 6206 and 6207 case, the bearing damage appeared to subside to a level similar to the Code 20 or 24 cases. The change in bearing damage may be the result of the particle impacts relative to the respective bearing size.

An additional speed case was considered similar to what was done with the Fatigue data. Figure 4.77 contains the average FFTs for every contamination stage for the different bearing sizes at the 1687 RPM case. Similar to the 1103 RPM case, the prominent bearing damage begins first in the higher frequency ranges due to the random particle impacts. There are still impacts in the

193
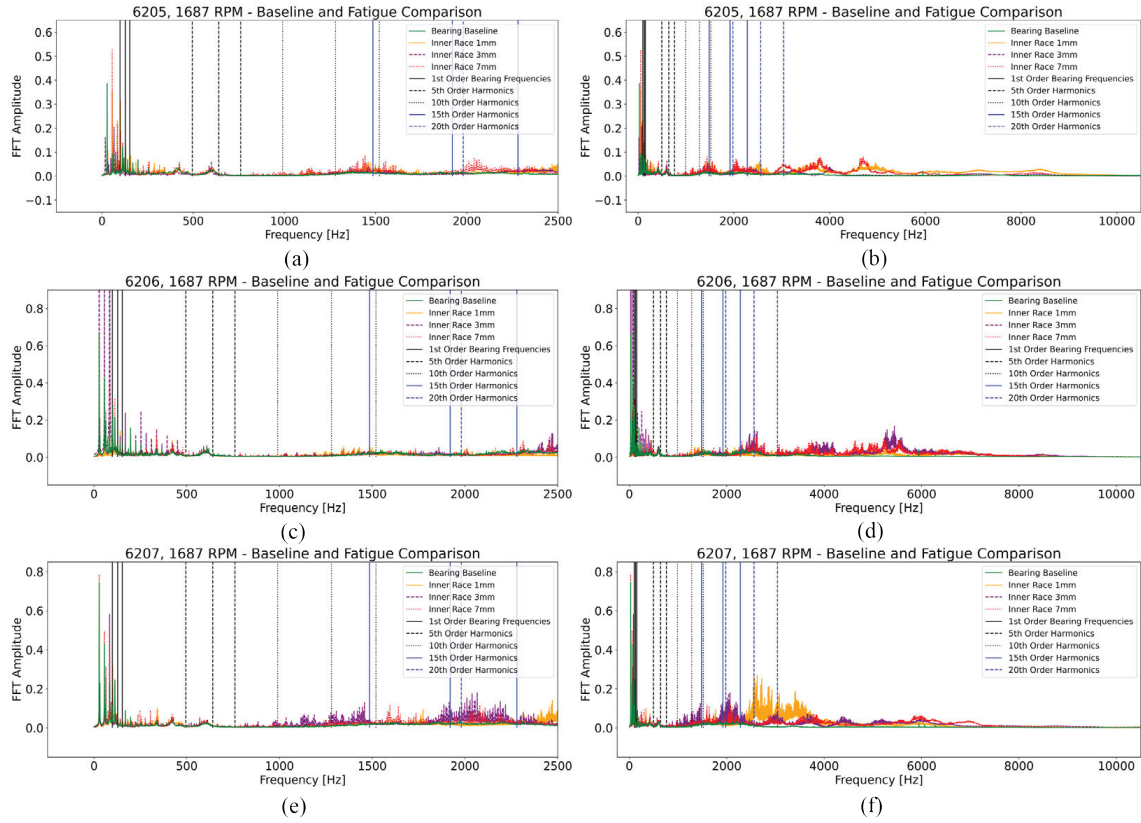
Figure 4.75: 1687 RPM case for the three different bearing sizes at the low frequency and full frequency ranges: (a) 6205 0 Hz - 2500 Hz, (b) 6205 0 Hz - 10000 Hz, (c) 6206 0 Hz - 2500 Hz, (d) 6206 0 Hz - 10000 Hz, (e) 6207 0 Hz - 2500 Hz, and (f) 6207 0 Hz - 10000 Hz

low-frequency range; however, they are muted in comparison and occur almost randomly. Hence, as the number of particles grows, such as in the 6207 case, the damage frequencies shift from the high frequencies to the low-frequency range (¿ 1000 Hz). Also, in contrast to the 1103 RPM case, each bearing Code 0 appears to be either the Code 20 or 24 case. The higher speed may cause an increased mixing of the grease and particles, increasing the number of impacts for the bearing case. Hence, removing the contamination particles reduces the impact in the high-frequency range.

The IMFs were investigated in addition to the average FFT representation to determine if there were specific bands that could identify certain defects over others. Figure 4.78 contains the IMF representations for the 6205 fatigue case as the bearing defects progress. Viewing the data across the different splits, the amount of overlap seen in the FFTs appears to carry over to the IMF case. Noticeably, the inner race 1mm case appears to exceed or resemble the 7mm defect case across the different bands. In IMF 3, the defect stages appear more interchangeable, whereas the

194
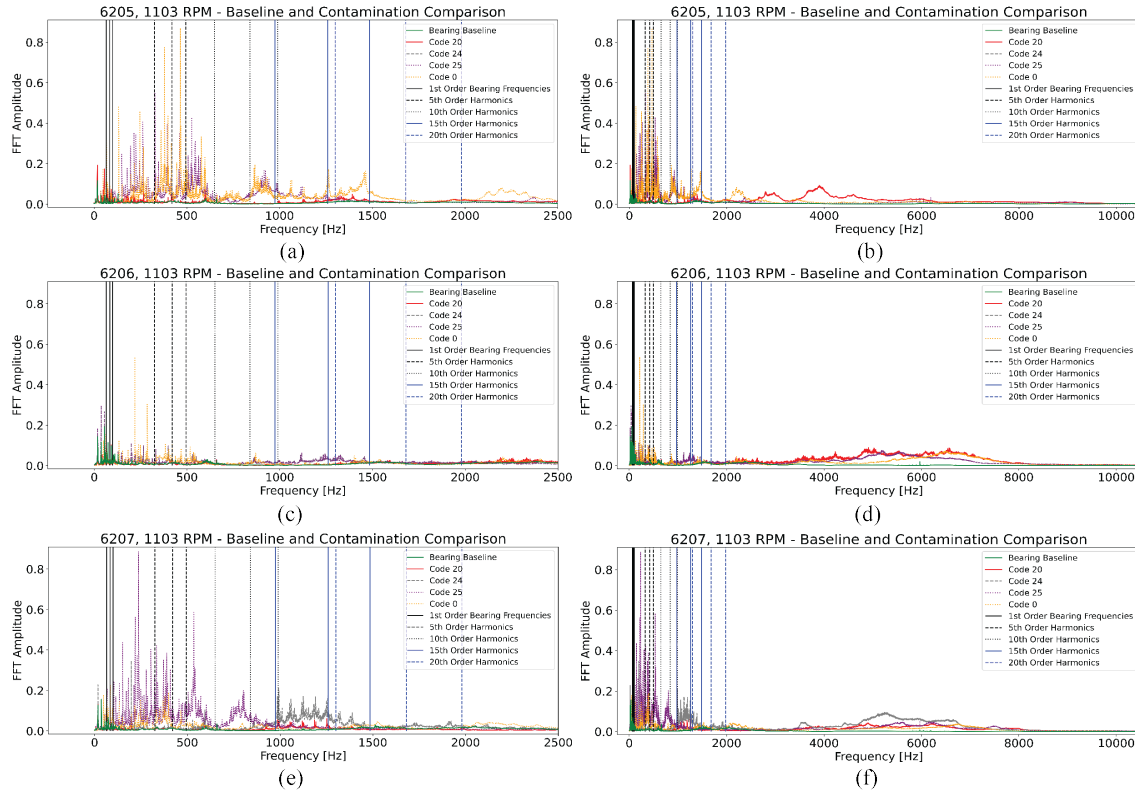
Figure 4.76: Average FFTs for the progressing contamination fault case for the 1103 RPM case: (a) 6205 0 Hz - 2500 Hz, (b) 6205 0 Hz - 10000 Hz, (c) 6206 0 Hz - 2500 Hz, (d) 6206 0 Hz - 10000 Hz, (e) 6207 0 Hz - 2500 Hz, and (f) 6207 0 Hz - 10000 Hz

baseline case appears to be significantly lower. However, there are instances where the baseline case is greater, such as in IMF 7 and successive cases. Hence, those impulse events related to the baseline case could be related to the speed and possible shaking in the test stand.

Figure 4.79 contains the IMF representations for the 6206 fatigue case across the different inner race defect cases. In contrast to the 6205 bearing there was an increased amount of noise/signals related to the inner race 3mm case. These observations occurred in the IMF 5, 6, and 7 cases. This could stem from additional noise or a defect in the bearing not anticipated from the fatigue case. In the IMF 5 and 6, it appears that the the baseline case is similar or greater than the the inner race defect cases, which should not be the case. Incidentally, the signals present in the IMF 1 and 2 case do establish a difference in the bearing cases. However, the progression does not appear as linear as thought. There does appear to be at least clear separation between the different bearing cases in at least several of the classes. For example, a higher signal denoted in the IMF 1 case, but not the

195
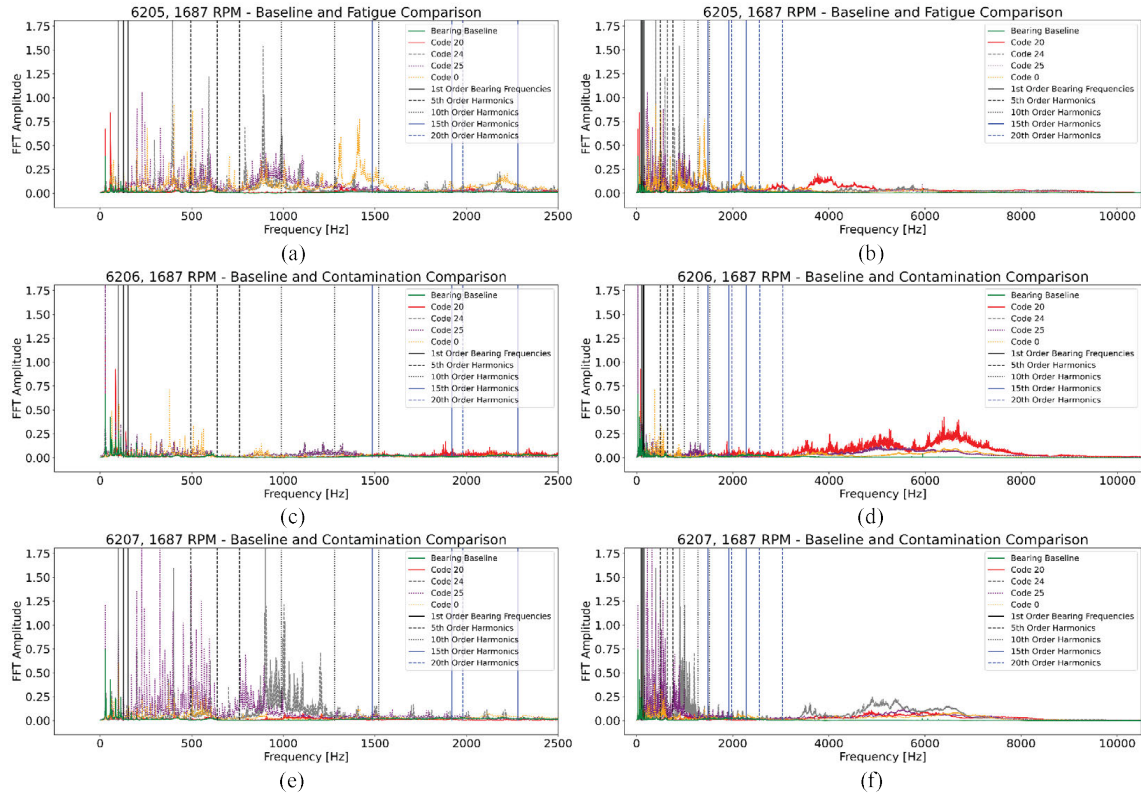
Figure 4.77: Average FFTs for the progressing contamination fault case for the 1687 RPM case: (a) 6205 0 Hz - 2500 Hz, (b) 6205 0 Hz - 10000 Hz, (c) 6206 0 Hz - 2500 Hz, (d) 6206 0 Hz - 10000 Hz, (e) 6207 0 Hz - 2500 Hz, and (f) 6207 0 Hz - 10000 Hz

IMF 5, 6, and 7 bands could signify an inner race 7 mm defect over a 3 mm defect case. It is still interesting to note the higher amplitudes in the lower frequency case over other cases.

Figure 4.80 contains the IMF representations for the 6207 fatigue case for the different inner race defect cases. For the 6207 bearing cases, the differences between the baseline and defect cases appear to occur in the higher frequency cases (IMFs 1, 2, and 3). As the defect appears to progress, the inner race 1mm case appears to have the highest amplitude in IMF 1. The 3mm defect case appears to have a higher amplitude in the IMF 2 case. Finally, the 7mm defect case appears to be greatest in the IMF 3 case. Outside of these three cases, the bearing data for the inner race defects appears similar for the different baseline defects outside of that defect case. In particular for the IMF 7 range, the inner race 1mm defect case appears lower in amplitude than the baseline case. It is possible that system noise would have caused the amplitude to be greater in this case.

From the fatigue case, there are distinctions in the class to class data. However, there are
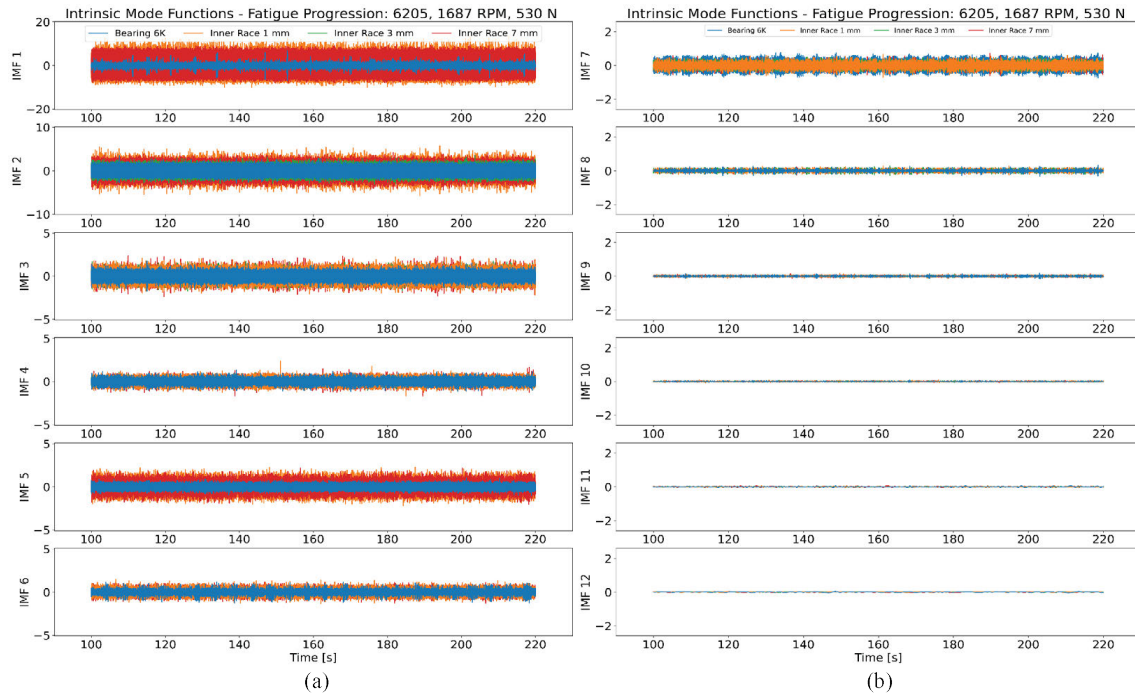
Figure 4.78: Fatigue Bearing data for the 6205 data split into IMFs (a) 1 - 6 and (b) 7 - 12
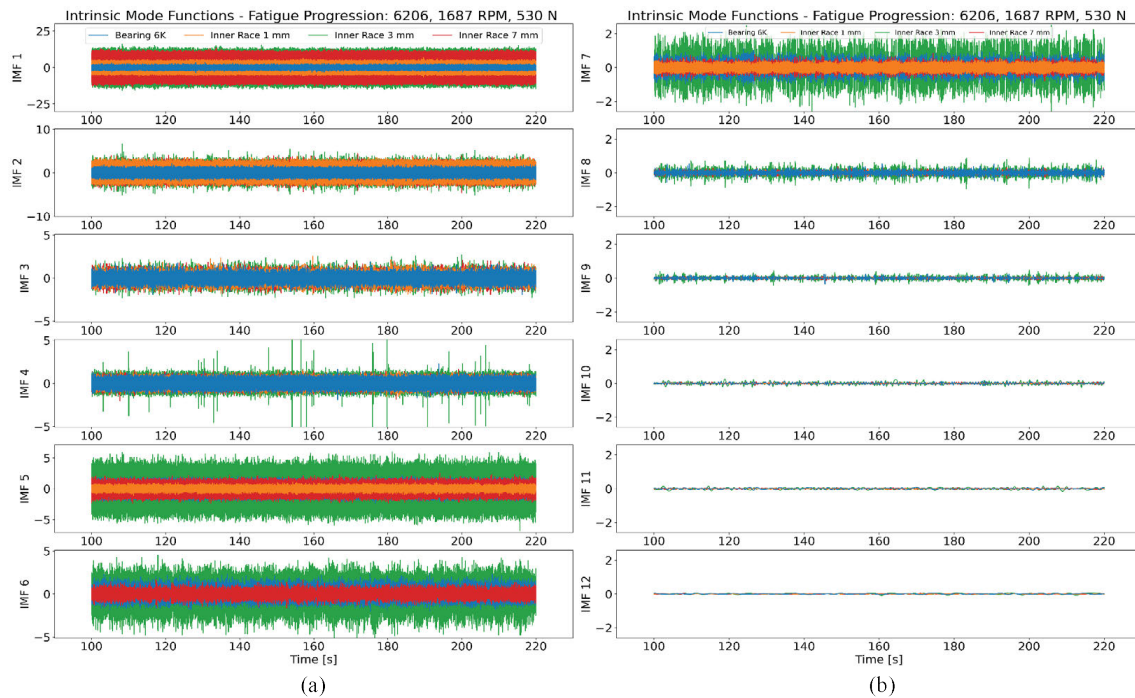
Figure 4.79: Fatigue Bearing data for the 6206 data split into IMFs (a) 1 - 6 and (b) 7 - 12
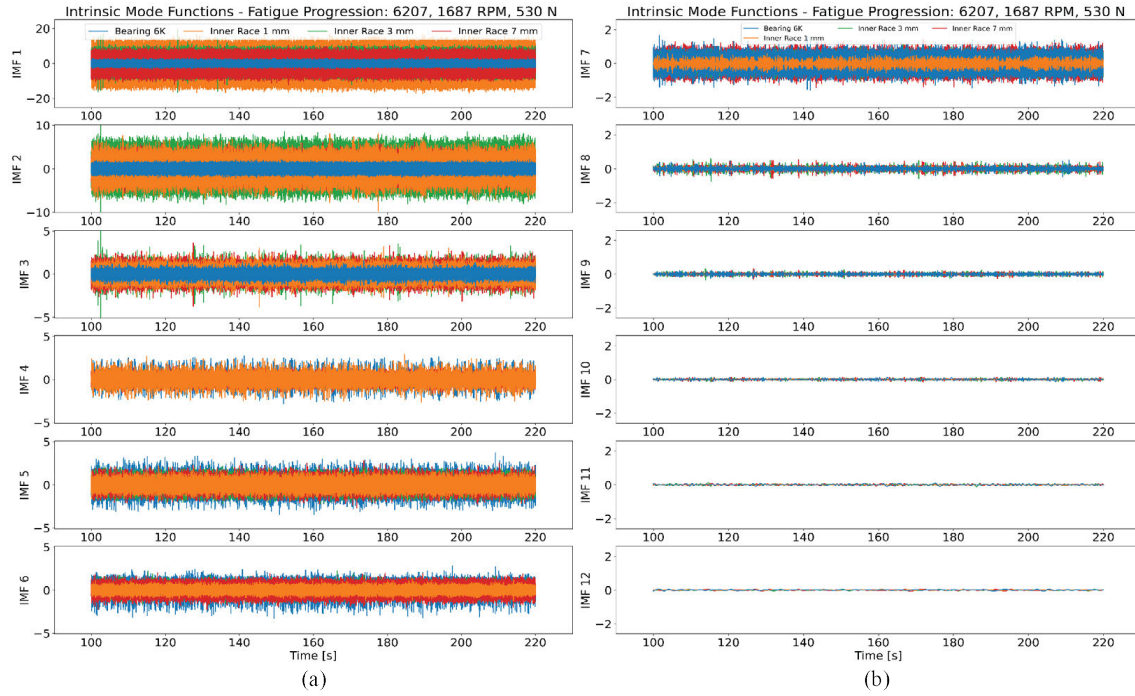
Figure 4.80: Fatigue Bearing data for the 6207 data split into IMFs (a) 1 - 6 and (b) 7 - 12

impurities in the data that could signify deviations to the normal trend in the data. The impurities appear to disrupt the anticipated defect progression for the bearing classes. The most likely cause for the deviations could stem from unaccounted elements in the test stand configuration. While that does lead to unanticipated errors in the lab system, these elements are accounted for and can be anticipated prior to the application of the data into a real world data.

In addition to the fatigue case, the contamination case was investigated to determine if there were similarities across the different bearing damage stages. Figure 4.81 contains the IMFs 1 through 6 and IMFs through 7 and 12 for the different contamination stages in the 6205 case. For the contamination data across the different bearing sizes, there appears to a large enough distinction in the bearing baseline data and the different contamination stages. For the 6205 bearing data, the only IMF that follows the expected progression clearly is IMF 2. However, in IMF 1 and IMF 3, the contamination cases for Code 24 and Code 25 appear similar enough to represent the deviations in the data. In IMF 4 and 6, the Code 24 data appears greatest, whereas the IMF 6 and 7 data appears to have the Code 20 case.

Figure 4.82 contains the IMFs 1 through 6 and IMFs through 7 and 12 for the different
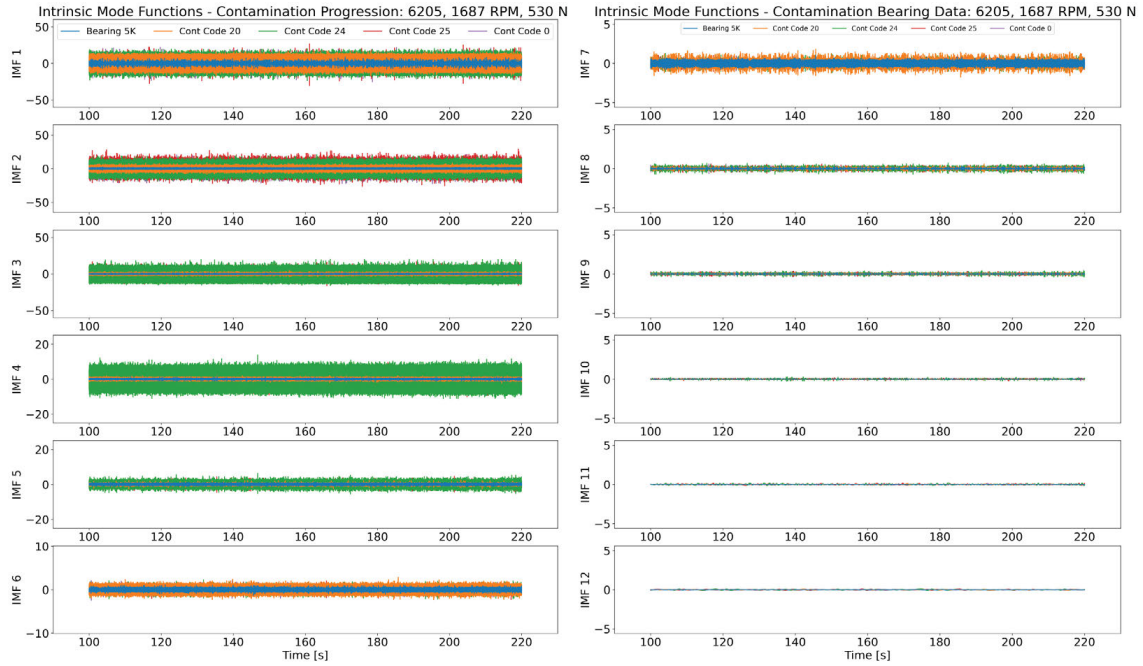
Figure 4.81: 6205 Bearing case for the progressive contamination case: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

contamination stages in the 6206 case. Similarly to the 6205 case, the contamination data appears to have different peaks in the IMF bands to signify different potential defect classes. However, it is still anomalous to see such a higher variation in the contamination Code 24 case in the IMF 6, 7, and 8 cases as compared to the other defect stages. It is possible that the deviation caused to that particular failure class is caused by the accumulation of material into clumps. The clumping of contamination material would reduce the number of possible continuous contact and cause possibly a similar signature to what is experienced in the fatigue case.

Figure 4.83 contains the IMFs 1 through 6 and IMFs through 7 and 12 for the different contamination stages in the 6207 case. In the 6207 contamination case, the progression appears to occur in a expected progression, where the Code 25 case appears to be the large amplitude condition in many of the expected failure mode cases outside of IMF 1. Following that observation, contamination Code 24 appears as the next highest amplitude. The distinction of the different IMF classes helps to signify the difference in the contamination classes between and also possibly where the impact of more versus less contamination.

From the contamination data, there is not as linear a line between the different contamination stages when considering the particles are captured in grease. A possible reason for the

199
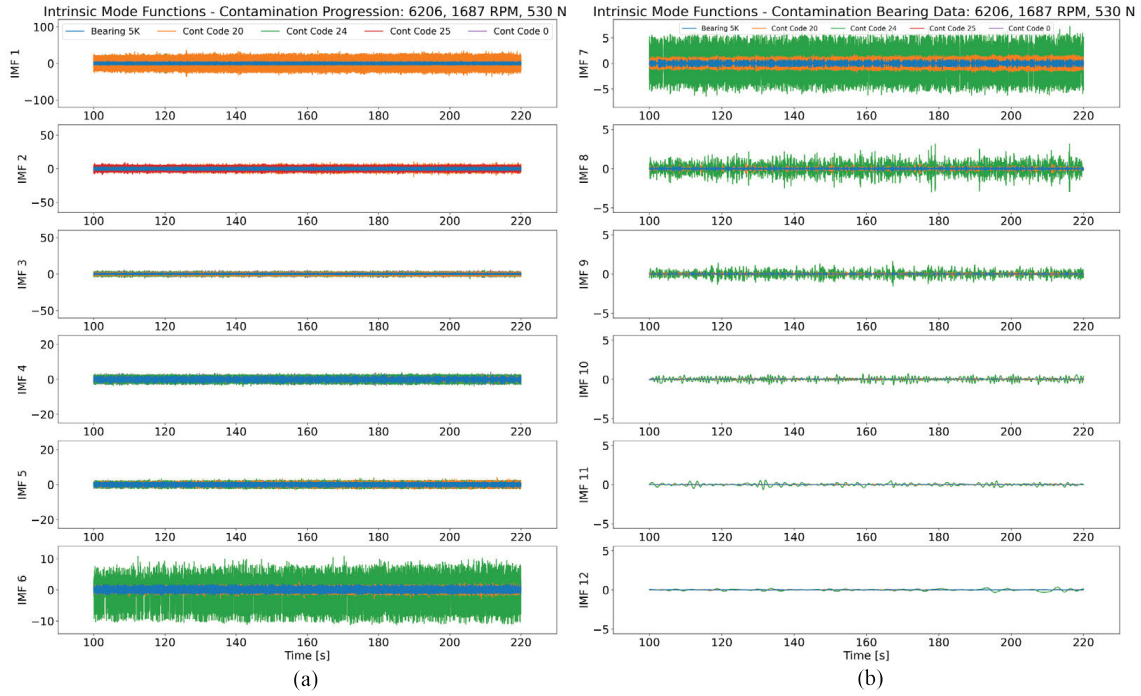
Figure 4.82: 6206 Bearing case for the progressive contamination case: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

occurrence could stem from the dispersion of particles through the grease as the bearing rolling elements travel around the bearing raceway. The clumping of the grease particles could cause higher vibration readings as the rolling elements pass over certain points in the bearing raceway. In each bearing size instance, though, the Code 0 bearing does not appear to come through the greater contamination stages. However, as evident from the average FFT stages, the Code 0 case does in fact appear higher than the average baseline case.

The progressive increase in the failure modes does signify, that there is an accumulation of damage as the defects progress. The progression does appear to differ depending on the failure mode induced. For the fatigue defect case, the bearing defects appear more gradual around the lower frequency whereas there is a marginal increase in the higher frequencies. In the contamination case, both frequency ranges appear to have a substantial increase relative to the baseline case. However, when removing the contamination particles, the upper frequency range appears to resemble the bearing baseline with a higher low frequency case. That could signify potential permanent damage left by the contamination particles. Hence, the continuous damage from the bearing defect and changes to the bearing surface roughness could cause changes to the overall bearing system in

200
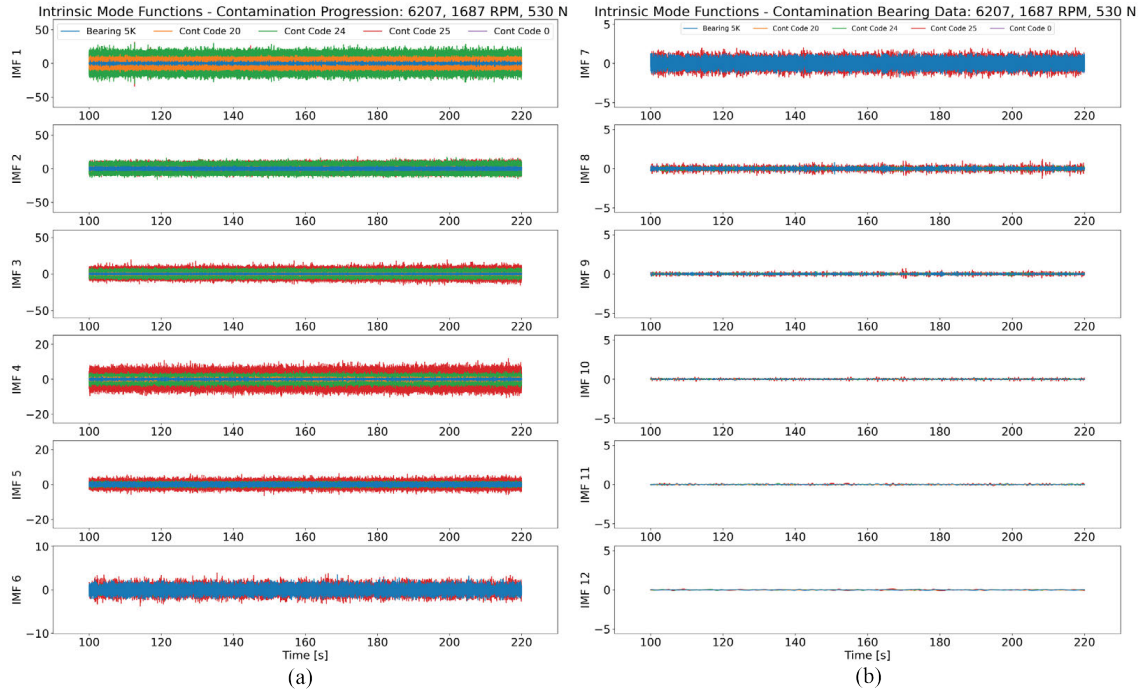
Figure 4.83: 6207 Bearing case for the progressive contamination case: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

the low frequencies. However, a permanent and substantial defect is needed to induce the higher frequency damage.

4.3.3 Load-Varying Failure Conditions

Load varying collections were added to determine the impact to the data as the load changed during collection. In industrial systems, bearings can experience higher loads as a result of a change in the system. An example of the load changing in the system is a CNC spindle cutting through material under different depths of cuts and material feeds and speeds. The force experienced on the tool and subsequently the spindle where the bearings are located causes changes in the vibration signal not directly related to the bearing speed. In the load-varying scenario, the test bearing is subjected to a 100 N increase in the force relative to the baseline bearing condition. In general, the amplitudes for the different IMF cases are expected to increase relative to the baseline cases for each bearing, due to the increase in force applied. The increase in force is expected to increase the rolling element deformation, which should increase the force acting on the bearing raceway.

Figure 4.84 shows the results of adding an additional load to the test bearing for the 6205

fatigue bearing case. In this instance, the main variation appears to occur in the IMF 5 case, which is where the bearing frequencies are primarily affected, and in the IMF 1 case. In this instance for IMF 5, though, the amplitude during the load varying stages is lower. In IMF 1 case, though, the amplitudes appear to increase, which could stem from an increase in higher frequency impacts from the rolling elements. For the defect progression, there appear to be different instances where the 1mm and 7 mm defect case are the higher class. However, the 3 mm defect case does not appear to come through, except in the IMF 2 case. The lack of differentiation between the different cases in a linear nature is unexpected and could stem from improper installation during the test phase.
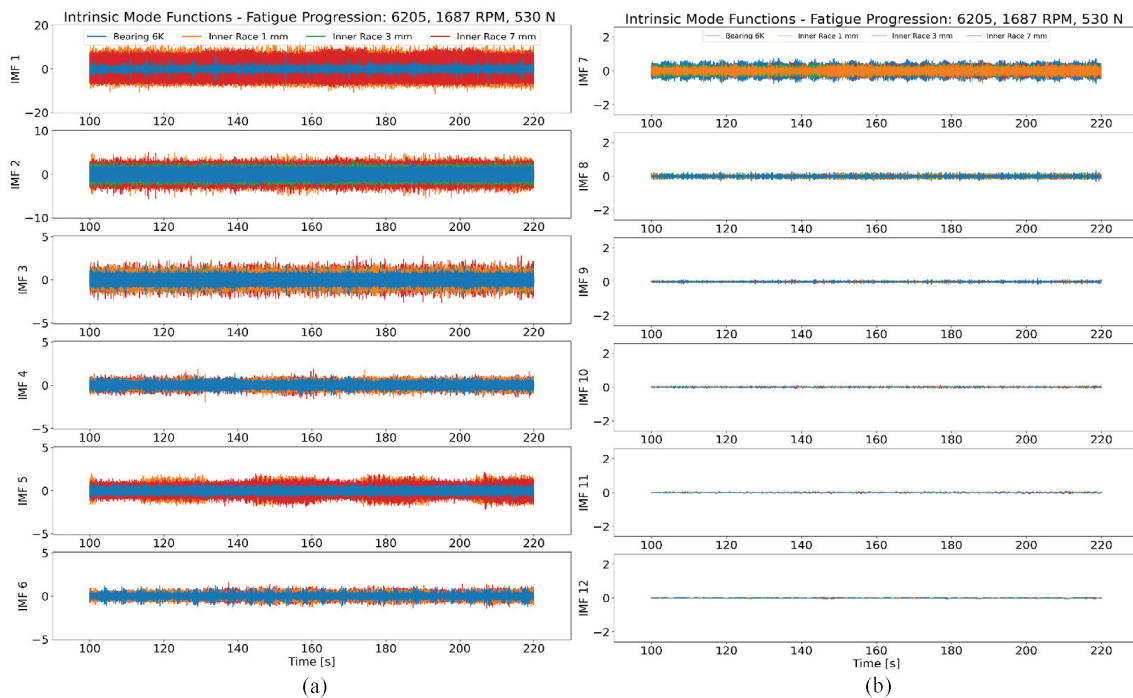
Figure 4.84: 6205 contamination bearing failure stage progression under changing load conditions: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

Figure 4.85 shows the results of adding an additional load to the test bearing for the 6206 fatigue bearing case. In the 6206 case, a similar occurrence happens with the 6206 data, that occurred with the 6205 data. As is evident from the 3mm defect cases, the IMF 1 through 4 had a higher amplitude, where as the IMF 5 through 7 cases had a lower amplitde as a result of the force applied. However, in the 1mm and 7mm defect cases, the bearing classes do not appear to increase at any stage except possibly in the IMF 3 case. A possible reason for the increase and decrease in different IMF bands could stem from the change in the frequency phenomena. It is theorized that

the 6205 case it could stem from the bearing impacts. In this instance, the 3mm defect case appears to validate this occurrence.
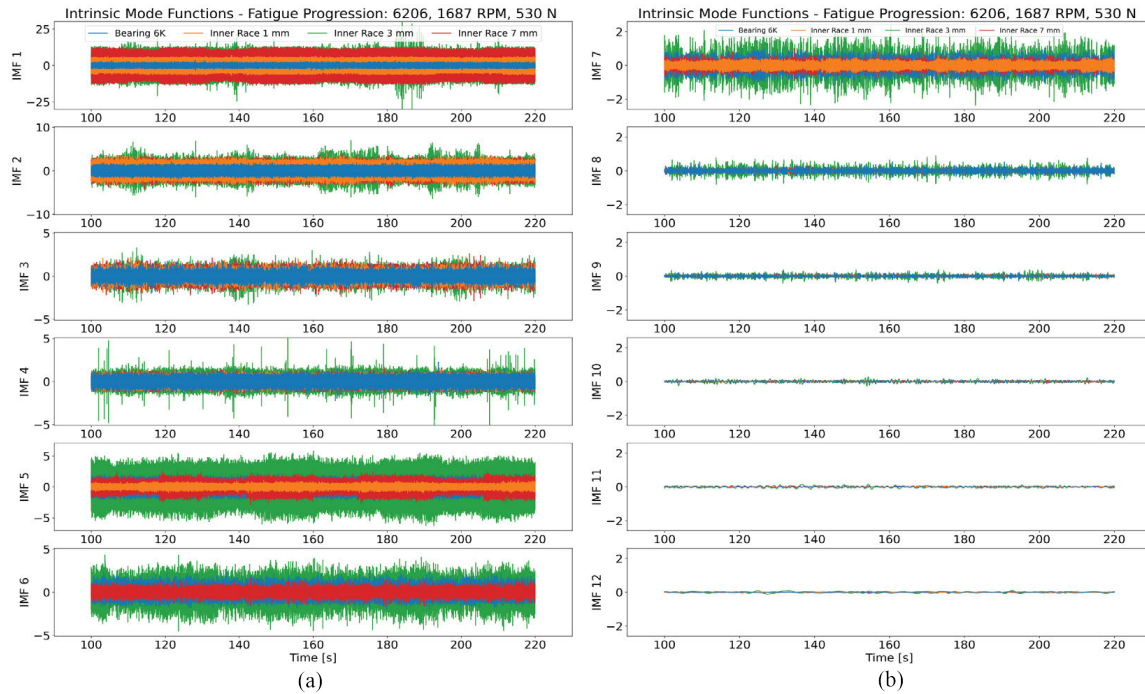
Figure 4.85: 6206 contamination bearing failure stage progression under changing load conditions: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

Figure 4.86 shows the results of adding an additional load to the test bearing for the 6207 fatigue bearing case. In this instance, the 1mm defect case did not appear to change except in the IMF 2 case. The other defect cases appear to follow a similar convention. In the 7 mm defect case, there is only a slight increase in the IMF 1 and 3 cases However, there are decreases in the IMF 6 and 5 case that is similar to the other bearing sizes. Unfortunately, a trend is not established based on observation for the different bearing defect cases through observation. While not trendable, it does appear that the most significant impact from the load-varying case involves an increase in the higher frequency amplitudes, but a decrease in the lower frequency amplitudes. It could stem from a dampening of the low frequency impact, but could increase the effects of the force on the overall system.

Outside of the fatigue cases, again, the contamination cases were tested. For the contamination cases, it is not as well understood what would happen in the contamination data as what would be expected in the fatigue data. Prior testing of the contamination particles [483, 503, 87]
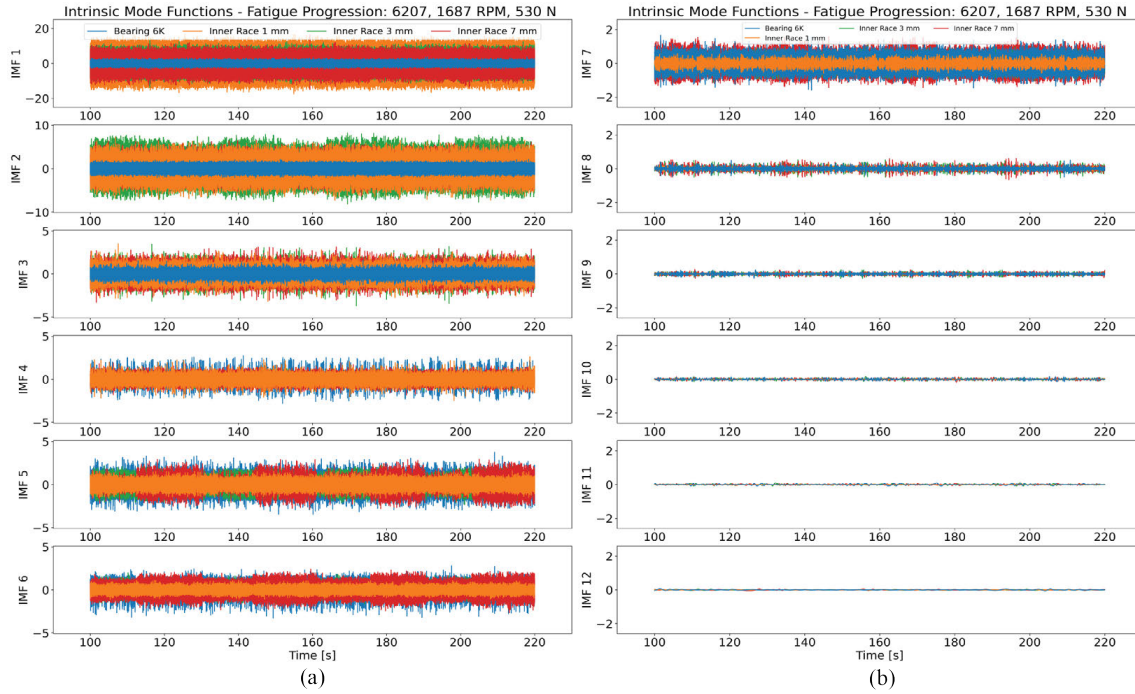
Figure 4.86: 6207 contamination bearing failure stage progression under changing load conditions: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

focused mainly on the effect of particle size and the speed effect. It is understood that the increase in speed causes additional impacts between the rolling elements, raceways, and particles. For the force impacts on the bearing, the increase in load may cause a sharper series of responses along the bearing surface and increase the defect size caused by overrolling the contaminated elements.

Figure 4.87 shows the results of adding load to the test bearing for the 6205 contamination bearing case. Unlike the fatigue case, the added force dampens the overall amplitude. It is possible that the added force dampens the impacts of the contamination particles with the bearing rolling elements. Hence, there is a lower impact on the bearing of higher frequencies. In addition, the amplitudes in the lower frequencies are also reduced as the load is applied. The effect could signify that the larger defect size from the fatigue case is impacted more by an increase in the load than in the contamination case. In the contamination case, the primary damage is formed due to the impact of particles on the rolling elements and raceways. Hence, when the force is increased, it may actually dampen the effect of those particles on the bearing raceway and relative rolling elements.

However, despite the conclusions and thoughts drawn from the contamination case, there are deviation in the 6206 and 6207 cases that support a different hypothesis. Figure 4.88 shows
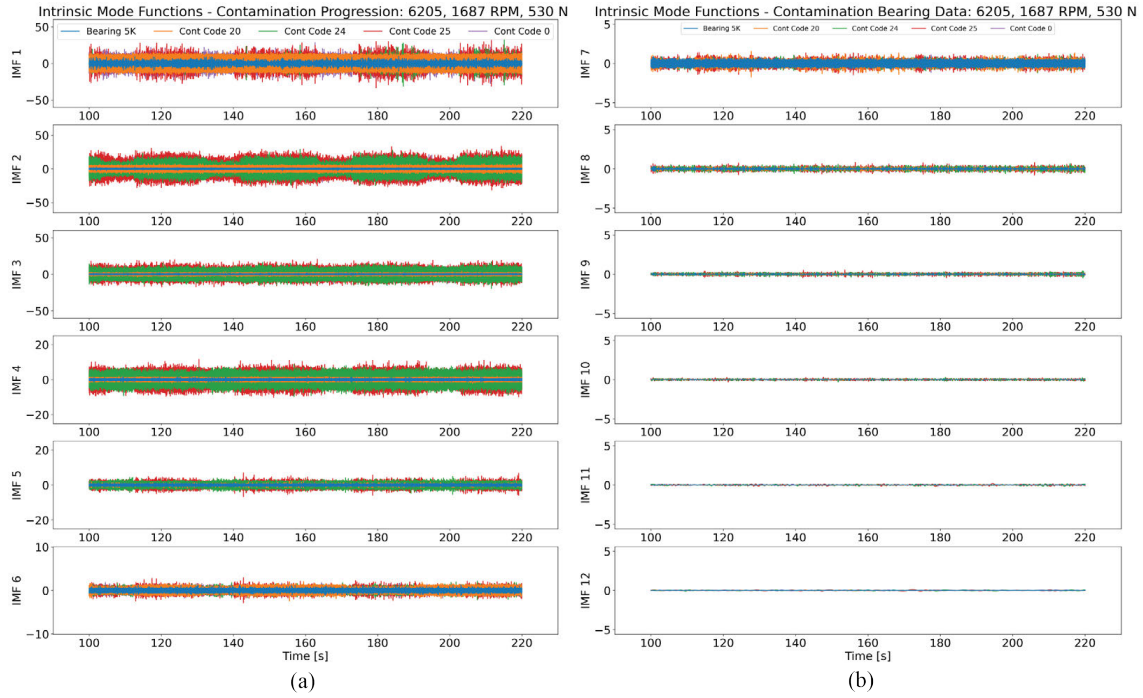
Figure 4.87: 6205 contamination bearing failure stage progression under changing load conditions: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

the results of adding load to the test bearing for the 6206 contamination bearing case during the test collection. The 6206 case is a deviation to the 6205 contamination case and the fatigue case defects by seeing an increase in the high frequency impacts and the low frequency impacts. In this defect progression, the amplitude for the contamination Code 24 and Code 25 cases increases in every instance for the different IMF band cases. The exception is the Code 25 case, which matches similar to what occurs for the Code 24 case. One possibility is that additional system noise affected the data that causing some data to appear over others in the different defect progression. Another possible impact could stem from the clumping of contaminated particles that could increase the signal response. Based on the Code 0 case, it would appear that there was some additional noise that was injected into the system caused by mechanical phenomena as that case increased relative to the contamination cases despite a lack of contaminated grease.

Figure 4.89 shows the results of adding load to the test bearing for the 6207 contamination bearing case. In the 6207 bearing case, the Code 24 case shows an increase in amplitude for the IMF 2 case, but a decrease in amplitude for the IMF 1 case. In the Code 0 case, there is an increase seen in the IMF 2, 3, and 4 cases, but a decrease is noted in the IMF 1 case. There is also a general
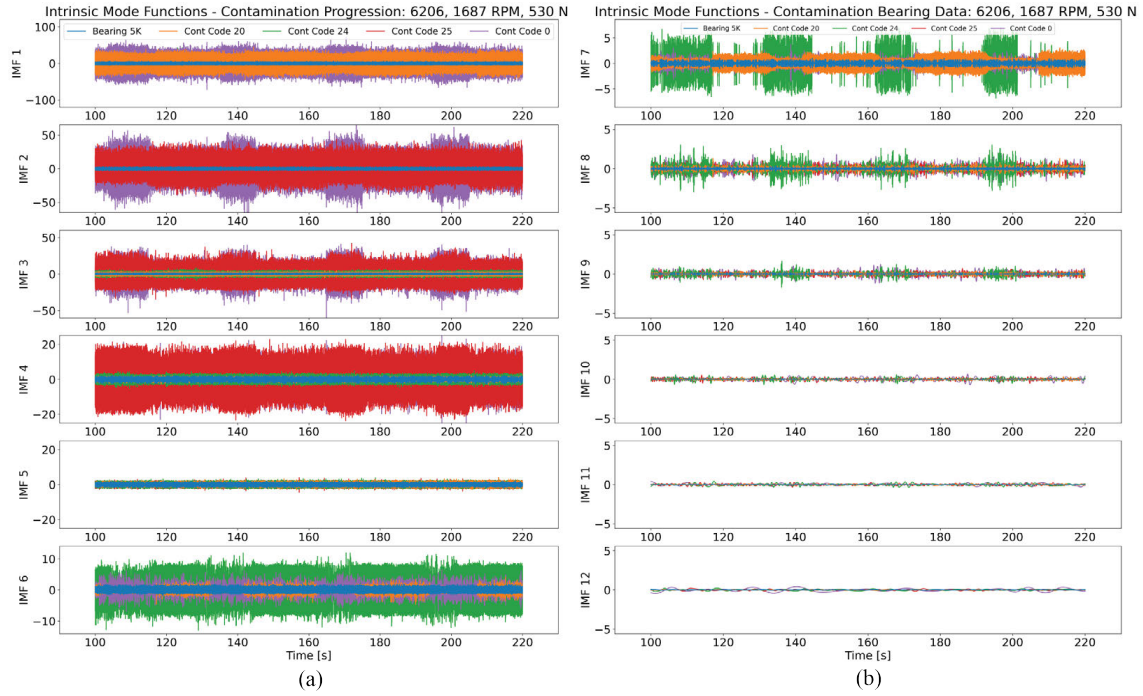
Figure 4.88: 6206 contamination bearing failure stage progression under changing load conditions: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

increase in the IMF 5 through 6 cases. For the Code 25 case, there is a slight increase in the bearing during the different loaded and unloaded moments for the IMF 4 case. While, there is a distinct change from the baseline case, the variation in the contamination cases could point to additional contamination not determined as a result of the system.

For the load varying case, it appears that for the fatigue and contamination case, different phenomena occur that signal a change in the defect state relative to the operating conditions. For the fatigue case, the bearing data appears to reduce in the lower frequency amplitudes but increase in the higher frequency amplitudes when an additional force is applied to the bearing. Certain defect stages appear to highlight this change over other defect stages, suggesting a level of noise that appears to affect the overall system. For the contamination stages, the effects appear more random. The effect of the particle size, the variability in the system noise, and the unpredictability in the particle effects appear to change the overall data effects. However, there is a clear differentiation between the changing load, suggesting that a characterization provided could help infer the contamination stage and a deviation from a baseline condition that is force-induced versus defect-induced.

Based on the observations from the IMF, the load-varying case appears to support different

206
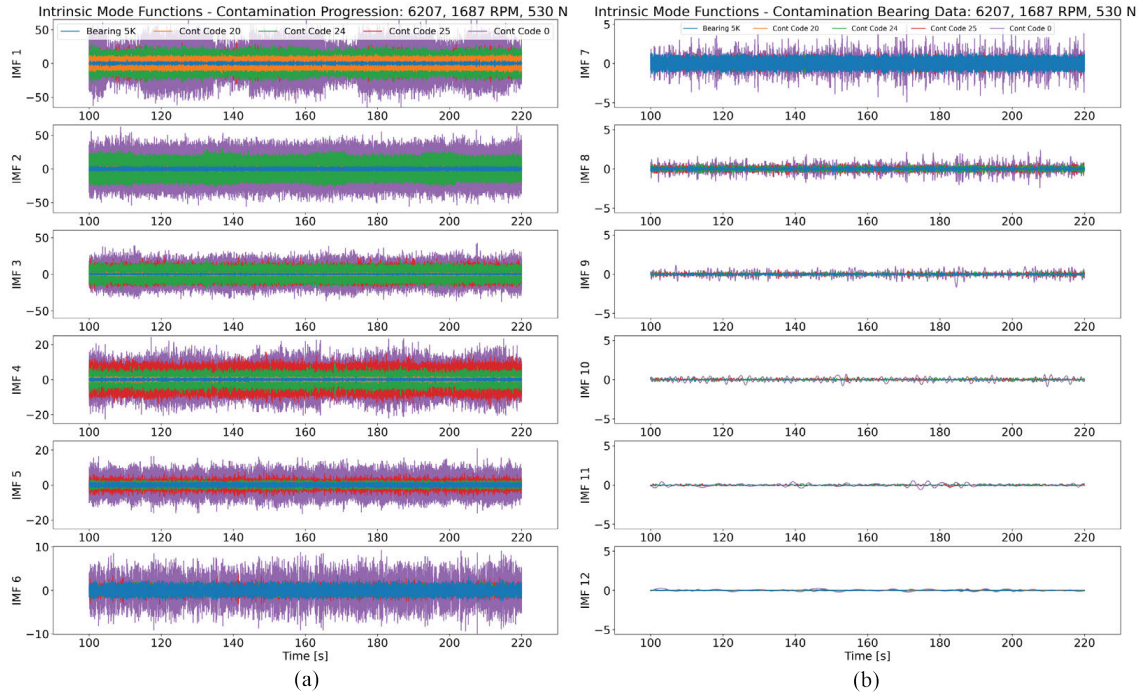
Figure 4.89: 6207 contamination bearing failure stage progression under changing load conditions: (a) IMFs 1 - 6 and (b) IMFs 7 - 12

conclusions based on the tested bearing failure mode. In the fatigue case, the bearing IMFs appear to increase with respect to the high frequency, but remain similar if not decrease in the low frequency cases. For the contamination case, it appears overall the bearing vibration decreases across the board. However, in the bearing IMF 2 and IMF 3 case, the bearing data appear to increase similar to the fatigue case. It is possible that the increase in force does increase the transmission of vibration, which would relate to an increase in frequency at certain modes. However, overall the force appears to more so dampen the vibration on the system at the lower frequencies. That could correspond to tigher connections of the overall system or even an unanticipated constraint in the system.

4.4 Chapter Conclusions

The overall goal was to create a methodology and generate condition monitoring data to formulate answers to the following three questions:

i.) Research Question 1.1: How are the existing bearing datasets deficient for training manufacturing condition monitoring systems?

ii.) Research Question 1.2: What methodology criteria are necessary to fill gaps for condition monitoring datasets?

iii.) Research Question 1.3: How much data are needed to train a bearing condition monitoring system for diverse failure modes?

For Question 1.1, there were a number (8) of different bearing datasets that detail varying methods of fatigue damage in different bearing configurations (*i.e.*, roller bearings vs. ball bearings, purposeful damage vs. run-to-failure damage, different failure modes). The breadth of bearing data is available; however, the dataset interpretations were not as solid overall. Equipment descriptions were a common basis between each different dataset. However, the damage level for comparison on a bearing-to-bearing case was lacking. For example, there were noticeable changes in the IMS bearing data to signify a changing health level. However, there is no relative reference to the actual "amount" of the bearing damage. Purposeful failure datasets incorporate some metrics related to damage size, such as in the CWRU and Paderborn datasets; however, these damage metrics are not necessarily considered standard.

Hence for Question 1.1, the datasets were assessed based on a set of standard criteria for datasets from Pipino *et al.* [2]. The existing condition monitoring datasets generally emphasized the following criteria: Appropriate Data Amount, Believability/ Free-of-Error, Completeness, and Consistent Representation/ Ease of Manipulation. Across the different bearing datasets, the primary differential factor on the failure modes considered and the level of Believability/ Free-of-Error. Another observation was that all the datasets focused on only one type of failure generation, whereas failures could occur under multiple failure conditions and scenarios. Hence, much of the condition monitoring research leveraging these datasets focuses on fatigue failures within bearings. However, other failures to consider are bearing assembly failures, contamination, lubrication damage and the data associated with these failures.

The brunt of the work considered for Question 1.2 was formulating a methodology that considered and linked back to existing design and condition monitoring approaches with these criteria in mind. The basis of the methodology considers an overall system where a component is critical to the operation and requires condition monitoring. The number of failure modes tested is based on a failure analysis conducted for the machine. In this frame of reference, the system is such that it contains multiple components that could be operating with the test component. The motivation

of the methodology, though, helps track the different methods of generating data, the formulation of the test system and the relation to the original system components, and the amount of data generated and labeled.

Hence, the data generated to answer Question 1.3 is based on the four different failure causes described by SKF [7] and their subsequent effects. In this research work, due to the criticality of these data features from industry, all of the major failure modes are considered to be tested. Three bearing sizes (6205, 6206, and 6207) were tested under different operating conditions: steady-state, speed-varying, and load-varying. In the characterization of data, the horizontal vibration data was considered due to the consistency of results found in the bearing baseline data. The vertical baseline bearing data appeared to fluctuate concerning the different testing locations and the subsequent equipment operating around the test stand. Each baseline and failed bearing were tested with three verification methods to spot change over the bearing states: statistical time features, frequency analysis, and Hilbert-Huang frequency (HHF) analysis. For the three methods considered, HHF provided the clearest representation of what frequency bands had the most differentiating attributes between the different bearing health states. For the contamination and lubrication data, the high-frequency bands (1000 Hz to 10000 Hz experienced the most change relative to the baseline case. Changes were found for the inner race defect cases between running speed frequencies (around 60 Hz) and higher frequency content (10000 Hz). For the bearing Brinelling case, most of the changes noticed in the data appeared in the lower frequency bands. In terms of failure modes, each of the major assigned failure modes based on historical knowledge should be tested as unique characteristics are noted in their signals that separate them from each other as distinct conditions. Research Question 1.3 discusses a secondary consideration regarding the physical amount of samples needed later to differentiate failure modes. However, the amount of data in diverse classes should consider the system's multiple failure cases, as evident from the different bearing failure frequencies and time requirements.

Hence, at this point, the following research questions could be considered answered based on the coloring. Green means completed and yellow provides a semi-completion. RQ1 and RQ2 provide a closed response through this chapter. Further investigation is conducted in Chapter 6 concerning RQ3. Hence, while an answer is provided in this chapter, further analysis is provided later to discuss the necessary physical data amount, leaving it yellow.

Formulation of a Digital Triple		
RA 1: Data Generation	RA 2: Data Transferability	RA 3: Data Analysis
RQ 1.1	RQ 2.1	RQ 3.1
RQ 1.2	RQ 2.2	RQ 3.2
RQ 1.3	RQ 2.3	RQ 3.3

Table 4.19: Current Status of the Research Goal as of Chapter 4

4.4.1 Limitations on Data Generation

One observation noted from the data generation is the amount of variance from the different data collections. As seen from the baseline results in Section 4.3.1, the baseline data differed for each collected bearing. There was similar frequency content between the different bearings tested; however, in the reflection of the time measures, the bearing data generated had a wide range of possible values influenced by the possible variations in the collected data. Some of the associated noise came from variations in peaks around the low-frequency content for the bearings. These variations are commonly associated with changes in the system alignment or the equipment installation. Due to the multiple individuals assisting in the data collection, variations could stem from the bearing installation and test configuration.

It should be noted that the variation in bearing vibration data is not necessarily an unexpected aspect. Similar to how image datasets have variations based on the focus, size of objects, and filtering used, vibration, in general, is subject to changes related to speed, load, temperature, lubrication, and overall geometry, to name a few parameters. Hence, as these parameters change unexpectedly, it is possible that the overall bearing vibration changes as a result, increasing the observed variance in the baseline data. Mitigating these quantities could stem from more rigorous control of the environment. However, it does allow resemblance to more of what is expected in an industrial scenario. The change in environmental and operational considerations could expect to see more of a long-term effect on the equipment versus short-term as was observed during this research.

A more difficult aspect of the data variance is that the damage implementation and progression results were also affected by the noise. As a result, the separation between the healthy and failure data is fuzzier due to the inclusion of noise. It could be possible to filter out these attributes through digital signal processing. However, subsequently, critical parts of the signal could, in turn, be missed. Hence, the variation in the damage data could affect the relationship of data from the surrogate triplet to the physical triplet for training data. The variance from any augmented data

may be further corrupted by incorporating a noise aspect. Hence, the variation in the damage data could affect the relationship of data from the surrogate triplet to the physical triplet for training data. The corresponding results may differ for the subsequent analysis of the data transfer methods described in Chapter 5 and the machine learning results in Chapter 6. It is possible in the future to consider a likelihood factor based on the equipment configuration to discard or keep anomalous data points. Incorporating that factor could, in turn, form the range of values needed for a particular training data set.

Chapter 5

The Digital Triplet: Data Transferability

After generating data from the surrogate triplet, additional knowledge is needed to determine how best to relate bearing data from a real surrogate system to resemble data from the physical system. In this chapter, the tools and methods investigated consider a physics-informed method using artificial vibration addition to known baseline signals, a technique we term as *Spectral Augmentation* for transferring data between different systems. A physics analysis of the bearing system is conducted to create a method whereby augmented data are filtered for feasibility and then scaled to levels appropriate for the real system. Section 5.1 describes the physics method, whereas Section 5.2 considers the physics-informed method. The following questions are considered in this chapter:

i.) Research Question 2.1: How is the artificially generated data from the surrogate system related to the production data taken from the real-world system?

ii.) Research Question 2.2: What considerations or knowledge are needed to relate different systems?

iii.) Research Question 2.3: Are the physical scaling and spectral augmentation sufficient for translating data from the laboratory environment to the real world when considering similar applications?

5.1 Physics-Informed Method: Physical Scaling

Physics scaling requires knowledge of the underlying system mechanics to formulate a physical model. Physical systems generally utilize a dimensional analysis model; however, these are heavily influenced by the surrounding system and applications. For this representation, a physics scaling method is created using the bearing test stand's Mass-Spring Damper systems (MSDs) to relate the generated bearing data between two different systems. In conducting this approximation, a physics normalization analysis is used to evaluate the effectiveness of this method in connecting bearing data between the different bearing sizes, operating conditions, and failure modes.

5.1.1 Physics Data Transfer

Similitude is a concept to verify the qualities between engineering models. The similitude concept is usually applied when relating an experimental system with a system acting in a real environment. An experimental system is said to have similitude with a real system when there is a geometric, kinematic, and dynamic similarity between two systems. The underlying study for similitude is traced back to Galileo's work [471] on the postulation of finding similarity between similar structures in different applications. Hence, the concept of similitude is generally tied to dimensional analysis.

Prior to real-world release, engineering applications go through different iterations of similitude testing to verify the performance of a product under real-world conditions prior to release. For example, wind tunnel testing is generally used to simulate real-world conditions on scaled-down models [472]. The similarity between multiple fluid mechanic applications stems from their relationships defined by Pascal's, Strouthal's, Reynold's, and other dimensionless numbers to ensure accurate condition modeling. Similitude analysis is also used for structures [504]. Prior to use in industry, structural materials undergo rigorous testing in the laboratory and scaled applications to ensure a complete understanding of the dynamic and fatigue loading limits. These materials are tested in scaled and full-sized models to verify any expected trends identified in the data. The similitude between two different applications is formulated based on a dimensional analysis.

The concept of dimensional analysis is to relate a system based on a number of physical quantities by identifying their base quantities. The formulation allows the representation of a n, the number of physics variables, to be rewritten into a set of equations to relate to the mechanical

system. The number of formulated dimensionless parameters is determined based on the n minus the rank of the dimensional matrix, m. The understanding of this relationship provides the basis for the Buckingham Pi Theorem. The mathematical representation of the Buckingham Pi Theorem is outlined in [505] and detailed in the following equations (Equations 5.1 through 5.3). Equation 5.1 considers a physical system comprising n physical variables. Equation 5.2 provides the reformulation by restating Equation 5.1 as a set of dimensionless parameters and equations. Each π term comprises a series of physics variables with exponential values attached.

$$f(q_1, q_2, ..., q_n) = 0 \tag{5.1}$$

$$F(\pi_1, \pi_2, ..., \pi_p) = 0 \tag{5.2}$$

$$\pi_i = q_1^{a_1} q_2^{a_2} ... q_n^{a_n} \tag{5.3}$$

For bearing condition monitoring, dimensional analysis is employed to verify vibration amplitude results for experimental and data-driven approximations. Jamadar *et al.* [506] considered 29 dimensionless parameters. Jamadar *et al.* [507] considered 20 dimensionless parameters in an additional paper. Jadhav *et al.* [157] considered 20 dimensionless parameters. Kumbhar *et al.* [508] considered 50-dimensional parameters in their investigation. The geometric, material properties, dynamic, and kinematic properties were considered in these parameters. Examples of these bearing and application properties are the bearing and shaft mass, the inner and outer diameter, the density, the shaft speed, the force acting on the bearing, the Modulus of elasticity, and bearing characteristic frequencies.

Outside of the dimensional analysis techniques, scaling laws describe the data transfer between a prototype system and a larger system. Scaling laws are generally categorized as an equation for understanding how vibration changes within a system. For rotor-bearing systems, these laws are generally coupled to similitude methods for understanding how speed and load affect the bearing condition. Li *et al.* [509] provided a partial similitude approach to determining a vibration scaling law for rotor systems as characteristics change dynamically. Luo *et al.* [510] demonstrated scaling laws to approximate the changes between rotor systems subjected to gravitational accelerations. In

both instances, the scaling laws primarily focused on changes to the system's critical speed and natural frequencies. This is similar to research conducted for structural analysis to determine how materials react under certain conditions with certain geometries in mind [511]. However, despite the research conducted on vibration scaling laws, it would appear that the common approach is still using dimensional analysis.

Combined with the formulation of bearing physics models discussed in Section 5.1.2, these parameters formulate the basis for modeling the bearing physics system. The dimensional analysis considered early has been tested to corroborate experimental validation results versus simulated results. However, these parameters have not been used to transfer data between bearings of different applications. Based on similitude analysis in other fields, such as dimensional analysis in fluid mechanics, it should be possible to apply a data-driven dimensional theory to relate bearing damage between different applications. Hence, for the physics data transfer, an MSD system returns an approximate vibration value of the expected bearing vibration data between bearings of different physical parameters. The approximate value is then used to scale bearing data for a source system (the surrogate triplet) to a target system (physical triplet) based on the two systems. The following subsection provides further background on the use of bearing physics and defect modeling.

5.1.1.1   Bearing Physical Models

McFadden *et al.* [302] proposed one of the initial bearing models to determine the vibration response of a bearing defect on the inner race of a bearing experiencing a constant radial load. The model considered bearing geometry, shaft speed, bearing loading, transfer function form, and decay. In providing these parameters, they determined the impulse of a rolling element rolling over the bearing defect and provided a representation of how it appears to decay over time. Equation 5.4 considers the impulses created by an inner race defect under a unit load from McFadden *et al.* To define nomenclature, d_0 is the impulse amplitude; the impulse function is represented by $\delta(t)$, where t is time, k is a constant, and T_d represents the time between impulses. T_d is the reciprocal of f_d, which is the rotational frequency of the defect for the bearing. The loading distribution is a function based on the bearing rolling element position, provided in Equation 5.5. The loading distribution outside of the bearing load zone is set as 0. The bearing loading considers the maximum load intensity (q_0, the load distribution factor (ϵ), angular point (θ), and a data-driven exponential (n). For ball bearings, the exponential is 1.5. Combined with impulse decay and vibration transfer, an

experimental representation could identify the impulse on the bearing as it rotates over time.

$$d(t) = d_0 \sum_{k=-\infty}^{\infty} \delta(t - kT_d)$$ (5.4)

$$q(\theta) = q_0[1 - (0.5\epsilon)(1 - \cos\theta)]^n$$ (5.5)

From the generalization of McFadden's model, other physics phenomena have been incorporated to capture the bearing representation. Lagrange's equation provides an example of capturing the bearing defect. As a brief recap of Lagrange's equation of motion, Equation 5.6 and 5.7 provide the basis and Lagrange's equations. The Lagrangian is the difference between the kinetic (T) and potential energy (V) of a system. q_i represents the coordinates of particles within the system, whereas $\dot{q}_i$ captures the velocities for these individual particles. Using these equations, it is possible to derive the generalized force acting on the bearing impulse defect.

$$L = T - V$$ (5.6)

$$\frac{d}{dt}\frac{dL}{d\dot{q}_i} - \frac{dL}{dq_i} = 0$$ (5.7)

Tandon *et al.* [335] provided the following derivations for the sine and cosine equations of motions for the outer race in Equation 5.8 and Equation 5.9 and the inner race in Equation 5.10 and Equation 5.11. q_i corresponds to the coordinates for the sine and cosine modes. Q is the generalized forces in the sine and cosine directions for the bearing. M represents the mass matrix for the bearing inner and outer race, respectively. The generalized force for the bearing rings is provided in Equation 5.12. X_i represents the mode shape used, p is the force as a result of the load on the rolling element and pulse shape. Subsequently, the radial load and mode shape change based on the defect location on either the inner ring, outer ring, or rolling element. The outer race defect remains stationary with respect to the bearing rotation, whereas the inner race defect changes location based on the rotation. The rolling element changes with the bearing rotation speed and the bearing axis. Figure 5.1 shows the rolling element load as a result of the rolling element position. Φ represents the rolling element position, whereas Φ_1 represents the limits of the bearing load from the rolling elements. Following this progression, Equation 5.13 provides the generalized format for the outer race defect

as a result of the mode shape, the rolling element load at a specific position, and the pulse shape.

$$\ddot{q}_{i,sin} + \omega_i^2 q_{i,sin}^i = \frac{Q_{i,sin}}{M_i} \tag{5.8}$$

$$\ddot{q}_{i,cos} + \omega_i^2 q_{i,cos}^i = \frac{Q_{i,cos}}{M_i} \tag{5.9}$$

$$\ddot{q}_{i,sin}' + \omega_i^2 q_{i,sin} = \frac{Q_{i,sin}'}{M_i'} \tag{5.10}$$

$$\ddot{q}_{i,cos}' + \omega_i^2 q_{i,cos} = \frac{Q_{i,cos}'}{M_i'} \tag{5.11}$$

$$Q_i = \int_0^{2*pi} p(\phi,t) X_i(\phi) \, d\phi \tag{5.12}$$
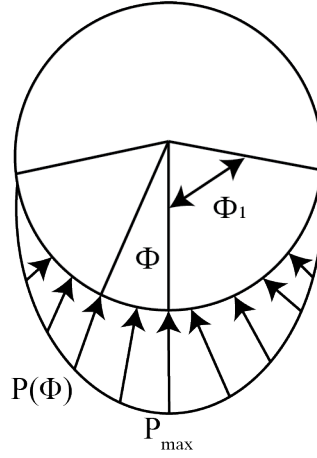
Figure 5.1: Change in bearing load as a result of rolling element position

$$Q_i = F(t) P(\varsigma) X_i(\varsigma) \tag{5.13}$$

From Tandon *et al.*, a 1-DoF model is created to estimate the vibration response from the bearing. The model DoFs increase based on the consideration of additional parameters. In Tandon *et al.* [?] considers only the y, denoting the vibration response relative to the force acting on the bearing from the rolling elements. Subsequent models have considered additional physical parameters to

217

consider different DoF models. Mufazzal *et al.* [114] considers a 2-DoF model considering the change in displacement between the shaft and housing. The displacement of the shaft and housing and the subsequent derivatives are represented by x_s and x_h, respectively. The supporting equation for the Mufazzal *et al.* model is provided in Equation 5.14. m_s and m_h considers the corresponding housing and shaft mass, respectively. c_s and c_e considers the shaft and lubrication damping, respectively. μ_i is the parameter to determine whether the rolling element is in the load zone. K_n represents the load-deflection constant based on the bearing geometry and material properties. Ψ represents the angular position of the rolling element relative to the shaft and housing displacement. F_{imp} considers the defect impact force and F_r represents the static equilibrium force to balance the total contact force due to the rolling elements in the load zone. Figure 5.2 considers the bearing model from Mufazzal *et al.* [114].

$$m_s\ddot{x}_s + (c_s + c_e)\dot{x}_s - c_e\dot{x}_h + k_s x_s + \Sigma_{i=z}^{Z}\mu_i K_{ni}(x_s - x_h)\cos\psi_i = 0$$

$$m_h\ddot{x}_h - c_e\dot{x}_s + (c_e + c_h)\dot{x}_h + k_h x_h - \Sigma_{i=z}^{Z}\mu_i K_{ni}(x_s - x_h)\cos\psi_i = F_{imp} - Fr$$

$$(5.14)$$
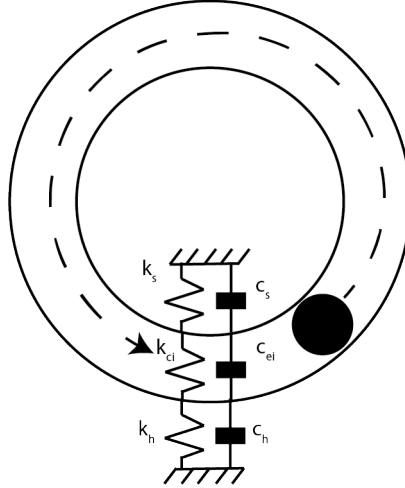
Figure 5.2: 2 DoF Model based on Equation 5.14

Sawalhi *et al.* [339] created a 5 DoF model that considered the displacement and subsequent derivatives in the x- and y-directions. The initial model was used in a subsequent representation containing 4 bearings, 2 gears, and two wheels. The general equations for the bearing system are represented in Equation 5.15. The bearing lubrication parameters are not used in this representation,

and only the rolling element stiffness is considered. In this representation, the housing displacement from Mufazzal *et al.* [114] is represented now as pedestal displacement in the x-direction as x_p and in the y-direction as y_p. The force function is calculated using the same force equation provided in Harris *et al.* [498]. In addition to these parameters, a resonance element is introduced to simulate the expected higher frequencies from the bearing system. Figure 5.3 considers the bearing representation from Equation 5.15.

$$m_s\ddot{x}_s + k_s(x_s) + f_x((x_s - x_p), \omega_s, dt) = 0$$

$$m_s\ddot{x}_p + k_s(y_s) + f_x((x_s - x_p), \omega_s, dt) = 0$$

$$m_p\ddot{y}_s + c_p\dot{x}_p + k_p\dot{x}_p - f_x((x_s - x_p), \omega_s, dt) = 0 \qquad (5.15)$$

$$m_p\ddot{y}_p + (c_p + cr)\dot{y}_p + (kp + kr)y_p - k_r y_b - c_r\dot{y}_b - f_y((y_s - y_p), \omega_s, dt) = 0$$

$$m_r\ddot{y}_b + k_s(x_s) + f_x((x_s - x_p), \omega_s, dt) = 0$$
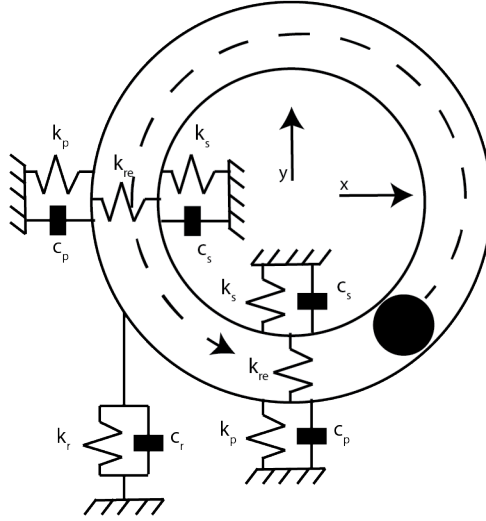
Figure 5.3: 5 DoF Model based on Equation 5.15

Lagrangian equations form the basis for most bearing modeling. Depending on the model, the number of parameters under consideration varies case-to-case. Rafsanjani *et al.* [512] considers the rotor and inner race of the bearing as a mass, where the damping is viscous and dependent on the rolling element stiffness and mass. Patel *et al.* [513] constructed a 3 DoF model that included the displacements of the rolling element along with the shaft and housing displacements. The model also

considered the mass of the rolling elements in the formulation. Nakhaeinejad *et al.* [514] utilized Lagrangian equations and bond graphs to formulate the model relationships for a 3 DoF model. This model considered similar relationships as considered in Mufazzal *et al.* [114], however in both the *x*- and *y*-direction, instead of only in the *y*-direction as was done by Mufazzal *et al.* [114]. Hence, a physical representation of a bearing depends based on the parameters and model structure used. Hence, the appropriate assumptions change on a case-to-case basis and are required in the formulation.

5.1.1.2   Bearing Defect Modeling

Despite the difference in bearing representations in terms of the parameters used and response, the bearing defect modeling remains consistent from model to model. The bearing defect forces are described using contact stress and the subsequent deformation through Hertzian stresses [338]. The bearing stress is modeled based on the point of contact shown in Figure 5.4. In this contact, two spherical objects act with respect to two planes. The bodies are the rolling element and the point of contact on the inner or outer raceway. From this point of contact, the bearing geometry is modeled based on the curvature of the two bodies. The sum and difference of the curvature are expressed in Equation 5.16. r_{ij} represents the radius of curvature of body i on plane j. ρ_{ij} is the curvature and the reciprocal of r_{ij}. The equations for deriving these particular elements are based on the geometry of the rolling elements and the bearing.

$$
\begin{aligned}
\Sigma\rho &= \frac{1}{r_11} + \frac{1}{r_12} + \frac{1}{r_21} + \frac{1}{r_22} \\
F(\rho) &= \frac{(\rho_{11} - \rho_{12}) + (\rho_{21} - \rho_{22})}{\Sigma\rho} \\
\rho &= \frac{1}{r}
\end{aligned}
\tag{5.16}
$$

From the bearing geometry model, it is possible to calculate the bearing deformation as either a point contact or line contact. A point contact is used for ball bearings, whereas a line contact is assumed for cylindrical rolling elements. Utilizing these relationships and the subsequent load on the bearing, it is possible to calculate the expected semimajor axis of projected contact in mm, a, the semiminor axis of projected contact ellipse in mm, b, and the resulting deformation, δ. The underlying assumption is that these are all steel bodies. Equation 5.17 provides the equations
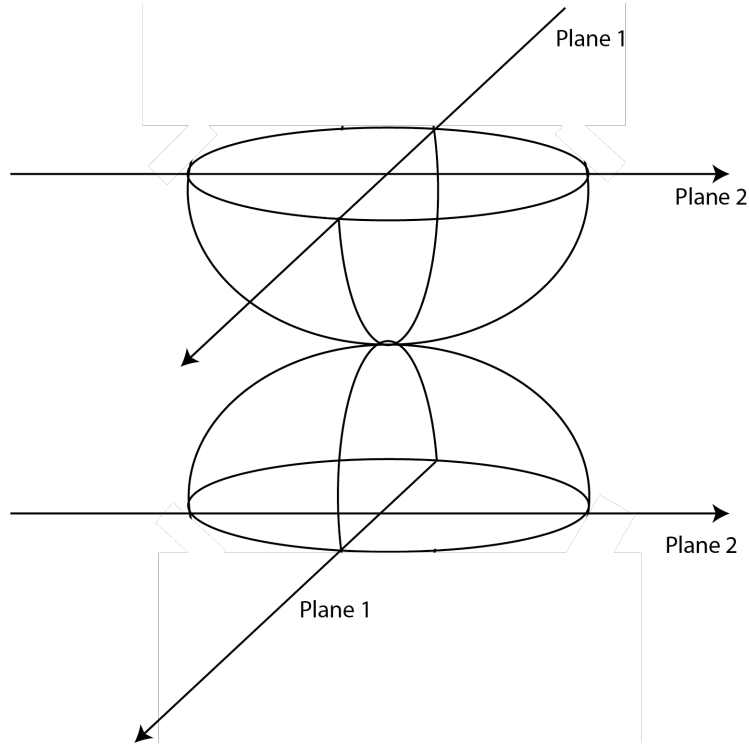
Figure 5.4: Bearing contact geometry

to calculate the deformation based on a radial load, Q, the bearing curvature sum, $\Sigma\rho$, and their respective dimensionless quantities. These dimensionless quantities have been derived by Brewe *et al.* [515]. A supporting table for these dimensionless parameters based on the curvature difference $(F(\rho))$ is provided in Table E.1. It was noted by Harris *et al.* that the bearing curvature difference for ball bearings is between 0.9 and 1 generally.

$$a = 0.0236a^*(\frac{Q}{\Sigma\rho})^{1/3}$$

$$b = 0.0236b^*(\frac{Q}{\Sigma\rho})^{1/3} \qquad (5.17)$$

$$\delta = 2.79 * 10^{-4}\delta^*Q^{2/3}\Sigma\rho^{1/3}$$

Subsequently, using these parameters, the area of contact for the bearing rolling element is approximated, and the resulting deformation in the contact area. Subsequently, the deformation can describe the load-deflection or load-deformation relation that the bearing rolling element acts on in a defect area. Harris *et al.* provides the following equation to approximate the load-deflection factor

221

for a ball-steel raceway contact in Equation 5.18. This equation was the basis for the figure provided earlier (Figure 4.25) in Section 4.3.1. Utilizing the load-deflection parameter, the rolling element force acting on a specific point is approximated using Equation 5.19. The deformation exponential factor is based on experimental testing with the bearing from Harris *et al.*. It changes to 10/9 for a line contact. K_p changes to K_l for a line contact bearing. Figure 5.5 provides a representation for how the deformation changes as a result of the applied radial force. In addition, the calculated load represents how much of the rolling element load acts on the bearing without considering the position in the load zone.

$$K_b = 2.15 * 10^5 \Sigma \rho^{-1/2} (\delta^*)^{-3/2} \tag{5.18}$$

$$Q = K_b \delta^{3/2} \tag{5.19}$$
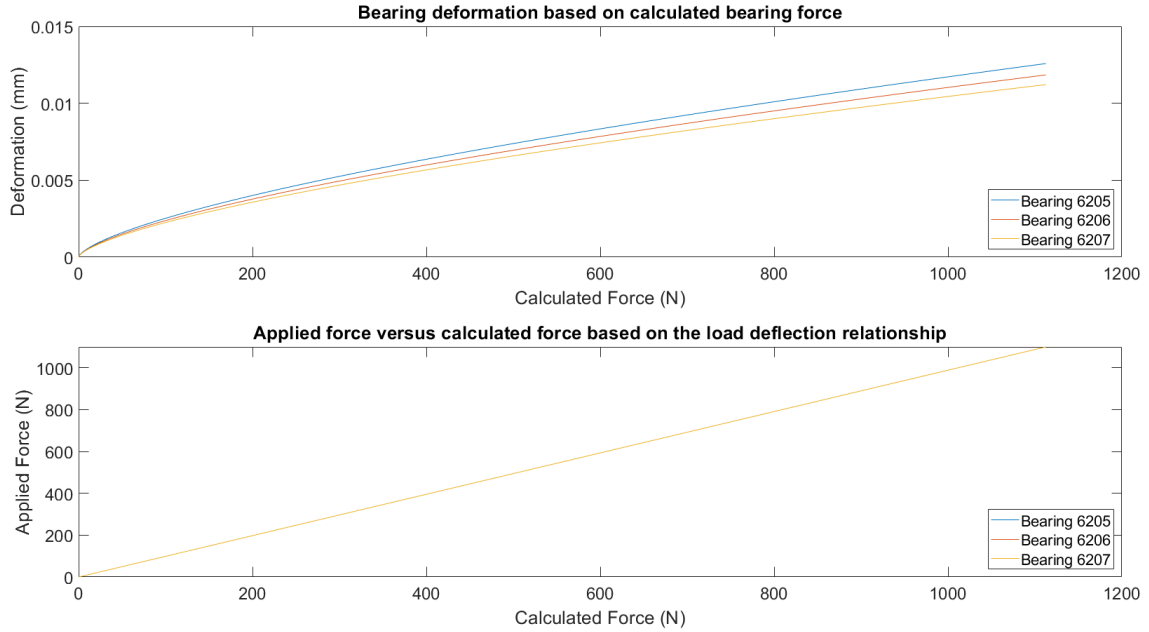
Figure 5.5: Bearing deformation and rolling element force as a product of the changing radial load on the bearing.

Until this point, the above description describes the bearing force at a predetermined location for the rolling bearing elements. The next aspect is to combine how the resulting force may change based on the introduction of a defect on the bearing raceway. The three cases considered in

prior literature are defects on the rolling element, inner, and outer race. The geometry and place-ment of the defect change in each instance based on the rotation of the rolling elements and the fixed bearing element. Further considerations are needed for the rolling element defect location as the defects roll with respect to the bearing shaft speed and then spin in place. Figure 5.6 shows examples of the geometry and kinematics associated with each bearing defect.
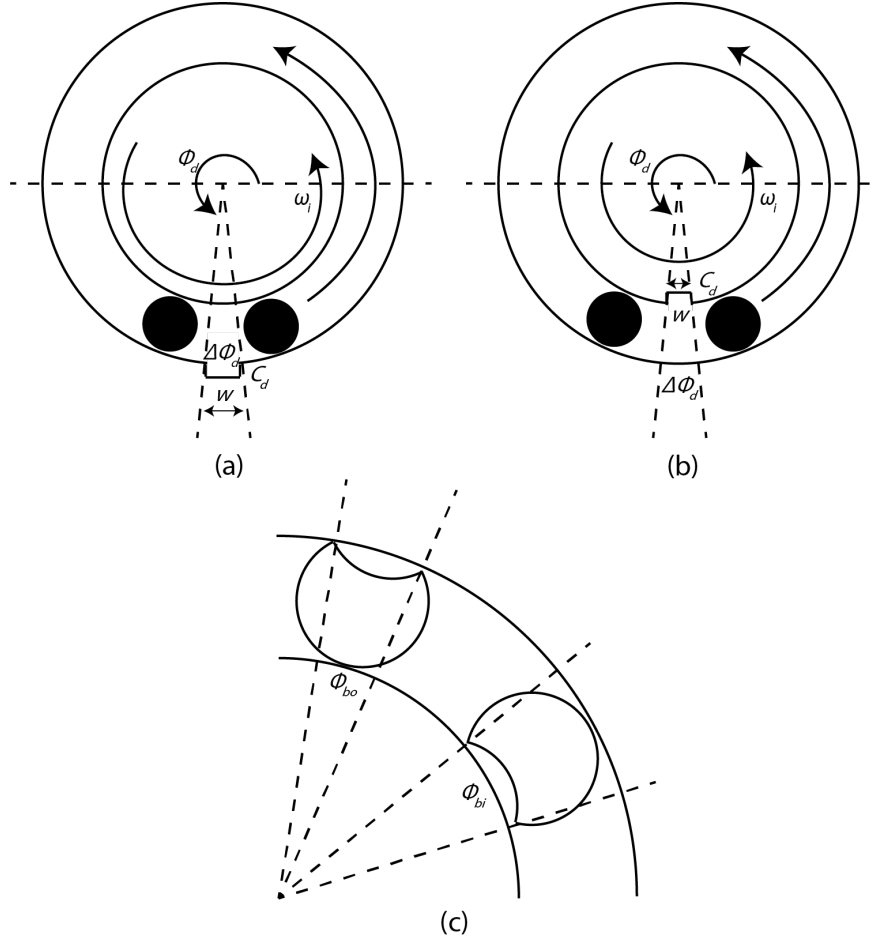
Figure 5.6: Bearing defect representation and modeling for the (a) inner race, (b) outer race, and (c) ball defect

Figure 5.6 (a) shows the outer race defect case, which remains stationary in each case. The defect is defined by a width of w. Depending on the bearing rolling element location, the width corresponds to a specific start and stop position. The total radial width of the defect is defined as two times the width divided by the ring diameter and is denoted as $\delta\phi_d$. The defect depth is defined by C_d. The bearing rolling elements then move based on the shaft speed with respect to

time. Figure 5.6 (b) shows the inner race fault condition. In this instance, the defect width location changes with respect to the inner race rotation and the bearing. The bearing inner race rotation is calculated in Equation 5.20. ϕ_0 is the initial defect location of the bearing on the inner race and ω_i is the inner race rotation with respect to time, t.

$$\phi_d = \omega_i t + \phi_0 \tag{5.20}$$

For the rolling element defect case in Figure 5.6, the defect width is defined relative to the rolling element as $\delta\phi_d$. For the bearing rings, δ_{bo} determines the radial defect for the rolling element in contact with the outer ring, whereas δ_{bi} determines the radial defect width with respect to the rolling element in contact with the inner ring. Since the bearing rolling element also rotates on its own axis with respect to the bearing, Equation 5.21 defines the ball spin speed relative to the bearing inner race speed. In equation 5.21, D is the bearing pitch diameter, and d is the bearing rolling element parameter. ω_i is the running speed for the equipment. The defect depth of the bearing is based on Equation 5.22. The maximum defect depth for when the rolling element is in contact with the inner race is C_{di} and is defined relative to the bearing inner ring diameter, D_i, and the center of the bearing width, x. The maximum defect depth for when the rolling element is in contact with the outer race is C_{do} and is defined relative to the bearing inner ring diameter, D_o, and the center of the bearing width, x. As the spall moves away from either raceway, the depth is defined as C_{dr} and is defined relative to the bearing rolling element diameter, d. The different defect depths are applied based on Equation 5.23.

$$\phi_s = \frac{\omega_i * D}{/} 2 * d(1 - (\frac{d}{D}\cos\phi)^2)t + \phi_0 \tag{5.21}$$

$$
\begin{aligned}
C_{di} &= \frac{1}{2(D_i - \sqrt{D_i^2 - 4x^2})} \\
C_{do} &= \frac{1}{2(D_o - \sqrt{D_o^2 - 4x^2})} \\
C_{dr} &= \frac{1}{2(d - \sqrt{d2 - 4x^2})}
\end{aligned}
\tag{5.22}
$$

$$C_d = \begin{cases} C_{dr} - C_{do} & if \ 0 < \phi_s < \phi_{bo} \\ C_{dr} + C_{di} & if \ \pi < \phi_s < \pi + \phi_{bi} \\ 0 & if \ otherwise \end{cases} \tag{5.23}$$

These defect locations are typically modeled based on the contact forces within a specific area, *i.e.* in the defect width. Equation 5.24 shows the deformation acting on the bearing rolling element without a defect. x_d and y_d is the bearing displacement relative to the bearing inner race and outer race. ϕ_j is the bearing rolling element position for the j rolling element. n_b is the number of rolling elements. c is the bearing diametrical clearance. The diametrical clearance for the bearing is taken from the bearing designation. There are five types of ISO bearing clearances: Group 2, Group N, Group 3, Group 4, and Group 5. Hence, the force equation in these system dynamic models is adapted to consider the forces acting on the defect location as a result of the Heaviside equation (shown in Equation 5.25). Hence, if the contact deformation of the bearing is positive, a resulting force acts on the bearing relative to whether the rolling element is in the load zone. Hence, the equation for the force becomes the following equation shown in Equation 5.26. In the instance where the rolling element hits the bearing in the defect width, the deformation equation changes to become Equation 5.27.

$$\delta_j = x_d \cos \phi_j + y_d \sin \phi_j - c, j = 1, 2, ..., n_b \tag{5.24}$$

$$\gamma_j = \begin{cases} 1 & if \ \delta_j > 0 \\ 0 & if \ \delta_j \leq 0 \end{cases} \tag{5.25}$$

$$f_x = k_b \Sigma_{j=1}^{n_b} \gamma_j \delta_j^{1.5} \cos \phi_j$$
$$f_y = k_b \Sigma_{j=1}^{n_b} \gamma_j \delta_j^{1.5} \cos \phi_j \tag{5.26}$$

$$\delta_j = x_d \cos \phi_j + y_d \sin \phi_j - c - C_d \tag{5.27}$$

Utilizing these equations with the bearing physical models, it is possible to approximate

the bearing defect response relative to the bearing physical parameters. Each scenario's bearing physical parameters are derived based on the application settings, primarily the load and speed, and the resulting assumptions related to the system. These assumptions relate to the shaft, housing, and bearing connections. These relationships are further investigated in Section E.1 to formulate the physics model. The following section, Section 5.1.1, discusses prior literature in how to leverage these models to relate bearing data between two different systems.

5.1.2 Bearing Physics Parameter Estimation: CGEC Test Stand

For the physics scaling process, the idea was to utilize the bearing physics model to scale data between the different bearing defect classes under similar and different operating conditions. In prior literature, the dimensional analysis methods have used parameters on a scale of 20 - 50 to describe the bearing system. In this analysis, 17 parameters are considered for the bearing physics model and are based on prior literature for modeling bearing vibration with defects. The process is not a traditional vibration scaling law as those do not seek to change the bearing vibration but more so to predict the potential outcomes. In this instance, the bearing data is modified to conform to the expectations of the bearing physics model to increase potential separability between the defect classes.

Three different bodies were considered in what would affect the vibration response from the system: the pedestal, the shaft, and the bearing. The pedestal refers to the bearing housing that clamps around the housing. The bearing shaft refers to the rotational element that is usually attached to the bore diameter. The third element considered is the bearing itself. Depending on the model, certain elements of these three bodies are used. For example, Sawahli *et al.* [339] considered all three bodies in their model formulation. Jiang *et al.* [178] considered only the bearing but split the model into three parts. The outer ring replaced the pedestal, and the inner ring was considered the shaft element. The grouping was justified based on the very small mass associated with the different bearing elements as opposed to the other system elements, such as the shaft and pedestal mass. The model bodies are combined into a 6DOF model is considered to capture the acceleration of the bearing under baseline and inner race defects. The model formulation is based off the Sawalhi *et al.* [339] derivation and Mishra *et al.* [516]. Table 5.1 contains the bearing model parameters for deriving the expected vibration limits.

The method for deriving the parameters varies between the different model formulations.

Table 5.1: Bearing model derived parameters

Parameters	Symbol	Unit	Description
Pedestal Stiffness	K_p	$\frac{N}{m}$	Stiffness for the housing holding the bearing in place
Pedestal Damping	R_p	$\frac{Ns}{m}$	Damping for the housing holding the bearing in place
Pedestal Mass	M_p	kg	Mass for the housing holding the bearing in place
Shaft Stiffness	K_s	$\frac{N}{m}$	Stiffness for the shaft the bearing rotates around
Shaft Damping	R_s	$\frac{Ns}{m}$	Damping for the shaft the bearing rotates around
Shaft Mass	M_s	kg	Mass for the shaft the bearing rotates around
Bearing Pitch Diameter	D_p	m	Average distance between the inner and outer bearing rings
Bearing Rolling Element Diameter	D_b	m	Bearing rolling element diameter
Rolling Element Number	N_b	-	Number of rolling elements the bearing has
Ball Stiffness	K_b	$\frac{N}{mm^{3/2}}$	Stiffness in bearing as it rotates around the bearing ring
Resonance Element Stiffness	K_r	$\frac{N}{m}$	resonance element stiffness in the bearing to excite vibrations
Resonance Element Damping	R_r	$\frac{Ns}{m}$	resonance element damping in the bearing to excite vibrations
Resonance Element Mass	M_r	kg	resonance element mass in the bearing to excite vibrations
Fault Depth	C_d	m	Depth of the bearing defect
Fault Width	w	m	Width along the length of the bearing raceway
Speed	ω_s	$\frac{rad}{sec}$	resonance element mass in the bearing to excite vibrations
Radial Force	Q_r	N	Force acting on the bearing surface

Essential Concepts of Bearing Technology by Harris *et al.* [498] provides a comprehensive review for deriving geometric, kinematic, and dynamic relationships in bearing systems. The resource was used to capture the loading and ball stiffness relationships. In addition, the reference included notes that were used to determine the contact deformation to describe the rolling element impacting the bearing surface. The contact captures the force acting on the bearing due to the radial load.

Assumptions were considered based on experimental measurements and prior literature. The mass terms are determined based on the weight of the objects using both a scale and calculated metrics based on the material properties. The material properties were taken from the material supplier, in this case, McMaster Carr, for parameters such as Young's Modulus and density. Three

different methods to consider for the stiffness parameters in calculating the stiffness for the block and shaft are considered in the research. The first two methods consider using ANSYS to solve for certain kinematic and/or dynamic parameters based on geometric models of the test stand. In one derivation, ANSYS determined the stress and strain for each nodal point of the FEA analysis to determine the material stiffness. This value was expected to correspond to the material stiffness from Young's Modulus. The second derivation considered the deformation at each point and determined the average stiffness by dividing the force by the deformation for each point. The initial form of the Ansys equation takes on the form shown in Equation 5.28, where K is the "global stiffness matrix" and F is the "global force matrix". However, the meaning of these terms changes based on what the problem is attempting to solve. For this particular problem, though, to solve for deformation (u) at each point, K represents the stiffness, and F represents the applied force at each point. For the stress and strain relationships, the K is the change in stress, whereas u is the change in strain for each point.

$$Ku = F \qquad (5.28)$$

The third method considers the use of beam stiffness calculations to approximate the stiffness for each element in the pedestal and shaft. Equation 5.29 is the equation for calculating the generalized stiffness. However, in the pedestal, the moment of inertia is subjective based on what moment is used. Hence, the minimum moment of inertia was considered and compared against other research papers and their results. Cao *et al.* [517], Kong *et al.* [163], and Mishra *et al.* [516] consider shaft stiffness ranges from in the *e7 and *e8 range for stiffness values. These values are dependent, though, on the shape of the system and could appear outside of that range. For ball stiffness, the bearing parameter is calculated using the equation from Harris book [498], as shown in Equation 5.18. The subsequent curvature summation and dimensionless deformation is calculated based on tables and equations from that book [498].

$$K = \frac{3EI}{L^3} \qquad (5.29)$$

The damping value was derived based on the stiffness and mass of each bearing. Sawahli *et al.* [339] approximated the damping values based on a percentage of stiffness. Damping for the shaft and pedestal are determined based on the Crandall [518] approximation of an ideal damper with

the Genta [519] incorporation for Hysteresis damping. Equation 5.30 provides the shaft equation for damping based on the interpretation from the Crandall interpretation of the loss factor of an ideal damper. η is the loss factor value based on the material, k is the shaft stiffness and ω_c is the shaft frequency. ω_c incorporates the whirling speed and the equipment speed. Jung *et al.* [520] provides a range of loss factors for steel between 0.002 and 0.02. These parameters are used to set the expected limits of the damping ranges.

$$r_s = \frac{\eta k}{\omega_c} \tag{5.30}$$

For the pedestal damping, ω_c should be considered the system natural frequency, ω_n. To calculate the natural frequency, Equation 5.31 is used, where k is the pedestal stiffness and m is the pedestal mass. The formula is based on second-order principles of system dynamics, formulated in Equation 5.31. x is the distance element, ξ is the damping element, and ω_n is the natural frequency. The second order principle provides the basis for the damping factor used by Genta *et al.*. The natural frequency for the pedestal is calculated as a result of the square root of the stiffness divided by the mass. The natural frequency is substituted into Equation 5.30 to derive the damping value.

$$\omega_n = \sqrt{\frac{k}{m}} \tag{5.31}$$

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = \omega_n^2 u(t) \tag{5.32}$$

The fault parameters are selected based on experimental testing. The parameters are labeled based on the fault depth and width, and it is assumed that the parameters are placed in the center of the bearing raceway. The depth is measured based on the control parameters of the CNC machine used to induce the defect on the bearing surface. The width is extrapolated using the tool depth and width and compared under a microscope. The bearing clearance factor is selected based on the bearing configuration. Bearing clearance depends on the bearing manufacturer; however, industry guidelines are based on ISO documentation [521]. Sawalhi *et al.* [339] configured the clearance level based on the preload interference. In this configuration, the clearance level is configured based on the stated bearing clearance from the manufacturer and the bearing bore diameter [1]. Harris *et al.*

---

[1] https://koyo.jtekt.co.jp/en/support/bearing-knowledge/10-2000.html

[498] does note that bearing clearance can potentially be negative, but that depends on the bearing lubrication as noted by Sawalhi *et al.*. The effects of bearing lubrication and the effects on parameters such as elastohydrodynamic lubrication of bearings are documented in works by Dwyer-Joyce [480] and Poddar *et al.* [86].

For the model used in this formulation, the original resonance parameters selected come from Sawalhi *et al.* [339]. The resonance parameters are tuned based on the expected response. In this case, the parameters were selected to represent a signal that would occur at 15 kHz. Similarly, these parameters could be modified or replicated to induce other frequency responses in the system that could simulate other bearing frequency responses.

5.1.2.1 Mass Parameters

Table 5.2 contains all of the mass elements used in the model formulation. The pedestal mass for the 6206 and 6207 bearings are similar due to the use of the same primary test blocks. Similarly, the mass shaft elements are similar due to the using the same base stock dimensions for the bearing dimensions. The bearing mass elements are based on the product information from their respective suppliers (Koyo, SKF, RBC, *etc.*). Since these manufacturers are held to ISO and ABMA guidelines, the mass should not change from manufacturer to manufacturer. The resonance mass element starts at 1 kg, but is further tuned based on the system output.

Table 5.2: Bearing mass elements

Element	6205 Bearing	6206 Bearing	6207 Bearing
M_p	4.96 kg	4.43 kg	4.49 kg
M_s	1.89 kg	2.26 kg	2.33 kg
M_b	0.13 kg	0.20 kg	0.29 kg

5.1.2.2 Stiffness Parameters

Table 5.3, Table 5.4, and Table 5.5 contain the different stiffness values calculated using each different method. The appropriate units accompany each entry. Table 5.3 contains the pedestal, shaft, and bearing stiffness values calculated using the stress-strain curve for the pedestal and shaft cases and Harris' equation for the ball stiffness case. The units for the ball stiffness case are kept as $\frac{N}{mm^{(3/2)}}$. The 1.5 factor is a constant due to the rolling elements being balls. Young's Modulus measures how easily a material can stretch or deform under different loads, which incorporates

the stiffness property of materials. In this comparison, the nodal points confirm that the inherent material stiffness is close to the expected material stiffness. For low-carbon steel, the material Young's Modulus is 2E11 $\frac{N}{m^2}$. The error between the simulation and the actual application is off by either +/- 5% for 4 out of 6 stiffness values. In addition, the stiffness is much higher than the expected approximation based on prior work from Sawalhi *et al.* and Kong *et al.*, where the pedestal stiffness is listed in the range of 1E7 $\frac{N}{m}$.

Table 5.3: Bearing stiffness elements based on ANSYS stress-strain curve and Harris *et al.*

Element	6205 Bearing	6206 Bearing	6207 Bearing
K_p	1.987e11 $\frac{N}{m}$	1.915e11 $\frac{N}{m}$	1.915e11 $\frac{N}{m}$
K_s	1.972e11 $\frac{N}{m}$	2.102e11 $\frac{N}{m}$	2.075e11 $\frac{N}{m}$
K_b	7.883e5 $\frac{N}{mm^{(3/2)}}$	8.631E5 $\frac{N}{mm^{(3/2)}}$	9.317E5 $\frac{N}{mm^{(3/2)}}$

The bearing values fall in the expected range of the stiffness values of the system based on prior literature. The formulation of these different parameters is purely based on the geometry; hence, the reason of why these parameters become larger is due to the increasing size of the bearings. In addition, if the load were to change, the bearing stiffness would change accordingly, whereas the bearing load increases, the stiffness increases, and vice versa. The reason for the larger stiffness is due to the larger contact area associated with the bearing.

Table 5.4 contains the stiffness values for the pedestal and shaft elements based on the ANSYS deformation calculations and calculation based on the force. In this approximation, the average stiffness value was an order of magnitude lower for the pedestal. In the case of the shaft stiffness, it was 5 orders of magnitude different. In this instance, the average was conducted after removing points that exceeded 90% of the maximum stiffness and were 110% greater than the minimum stiffness. The goal was to try and remove any particular outlier stiffness values in either case to consider only the average values. However, the stiffness values are still much higher than the expected stiffness values in the pedestal case.

Table 5.4: Bearing pedestal and shaft elements based on force and deformation relationships in ANSYS

Element	6205 Bearing	6206 Bearing	6207 Bearing
K_p	4.920e10 $\frac{N}{m}$	6.954e10 $\frac{N}{m}$	6.954e10 $\frac{N}{m}$
K_s	4.442e6 $\frac{N}{m}$	4.398e6 $\frac{N}{m}$	4.122e6 $\frac{N}{m}$

A final consideration was to conduct a generalized approximation based on the beam stiffness equations. The moments of inertia (I_{xx}, I_{yy}) were considered in two different orientations for the

pedestal block. The shaft stiffness was considered based on the moment of inertia. Table 5.5 considers the minimum values for the different moments of inertia. The parameters were within the expected range of the bearing parameters; however, they differ greatly from the ANSYS results. A possible reason is an over definition of the parameters within the ANSYS environment. The initial resonance element for stiffness is kept the same as with Sawalhi *et al.* is 8.8826e9 $\frac{N}{m}$.

Table 5.5: Bearing pedestal and shaft stiffness approximated based on calculated beam stiffness

Element	6205 Bearing	6206 Bearing	6207 Bearing
K_p	5.45e7 $\frac{N}{m}$	3.88e7 $\frac{N}{m}$	3.88e7 $\frac{N}{m}$
K_s	1.06E6 $\frac{N}{m}$	1.16E6 $\frac{N}{m}$	1.18e6 $\frac{N}{m}$

5.1.2.3 Damping Parameters

The damping parameters primarily depend on the inputs from the stiffness and mass inputs. Table 5.6 considers the bearings using the stiffness values from Table 5.3 and Table 5.6 shows the damping values from Table 5.5. In each case, hysteresis damping was considered due to the high stiffness value associated with the ANSYS model. The incorporation of hysteresis damping contributes to the potentially low quantities seen in Table 5.6. The range in damping values appears to cover the potential range seen in Sawalhi *et al.* and Kong *et al.*, who ultimately used a percentage versus an actual calculation to determine the damping value. The lower damping values also have a damping ratio that allows them to consider the expected range of what is found in bearings based on damping testing conducted by SKF.

Table 5.6: Ideal Damping with hysteresis damping using Table 5.3 stiffness values at 1687 RPM

Element	6205 Bearing	6206 Bearing	6207 Bearing
R_p	1.985e5 $\frac{Ns}{m}$	1.842e5 $\frac{Ns}{m}$	1.842e5 $\frac{Ns}{m}$
R_s	1.221e5 $\frac{Ns}{m}$	1.378e5 $\frac{Ns}{m}$	1.390e5 $\frac{Ns}{m}$

Table 5.7: Ideal Damping with hysteresis damping using Table 5.5 stiffness values at 1687 RPM and considering a damping ratio factor of 2 and 2.5 for the pedestal and shaft respectively

Element	6205 Bearing	6206 Bearing	6207 Bearing
R_p	331.64 $\frac{Ns}{m}$	304.92 $\frac{Ns}{m}$	264.72 $\frac{Ns}{m}$
R_s	29.41 $\frac{Ns}{m}$	30.11 $\frac{Ns}{m}$	30.2425 $\frac{Ns}{m}$

5.1.3 Bearing Physics Parameter Estimation: Real World System

A bearing system in the real world was theorized on what the parameters were based on an understanding of the formulation of these parameters in the test system, as shown in Table 5.8. However, the theoretical values would require additional considerations that are not readily available. As an example, a large mechanical system was considered, in this instance, a motor from ABB[2]. Only the pedestal mass was approximated based on information from the catalog. The bearing stiffness was approximated based on the bearing size associated with a range of potential shaft diameters using the Equations from Harris *et al.* [498]. However, it was realized during the formulation of the final parameters for the test system in the previous section that the system parameters can widely vary based on the assumptions considered in the base analysis. Therefore, any approximations of the production system in the real world required knowledge of certain mechanical aspects that were not readily available based on documentation from publicly available sources.

Table 5.8: Model elements from a 500 hp motor with a 6222 bearing

Model Element	Approximation Reason	Value
M_p	Based on ABB motor catalog	2150 kg
M_s	Approximated based on the weight of the rotor	?
K_p	Approximated based on the force, moment of inertia, and Young's Modulus	?
K_s	Approximated based on the force, moment of inertia, and Young's Modulus	?
K_b	Based on the bearing size and the dimensionless deformation contact	1.51e6 $\frac{N}{mm^{3/2}}$
R_p	Approximated based on the system run speed and the pedestal stiffness	Required knowledge from the pedestal stiffness
R_s	Approximated based on the system run speed and the shaft stiffness	Required knowledge from the shaft stiffness

Hence, certain elements regarding the mechanical system are still needed based on the overall design to utilize the physical scaling process. For example, the material for the motor frame is iron, which has a similar Young's modulus to steel. Hence, this could be foreseen as the maximum stiffness in the system, similar to the test system. The stiffness of the shaft is subjective, though,

---

[2]Accessed from https://new.abb.com/products/7BEM50504L-2340/em50504l-2340 on 10/23/2023

based on the material that makes up the shaft. In motor systems, the shaft materials are made up of steel with chromium, nickel, and molybdenum mixed in [522], causing a change in the material stiffness. In addition to the difference in the material properties, the change in shaft size further causes a change in the material stiffness. Hence, the motor geometric dimensions additionally affect the system configuration. Once these parameters are set, the damping ratios are configurable.

The lack of these parameters highlights a drawback of the physical scaling process and why most data-driven models are favored over their physical counterparts. The time to track down the information for a real world system could take weeks/months, if not a priority item for a product/support team. In the time it could take to train and deploy a physics model, a data-driven model could be trained and retrained hundreds of times over newly collected data and insight gained from the machine. However, that depends on the fact that the collected system data contains all the possible classes the maintenance staff wants to monitor and accurately represent the system class. Therefore, the data-driven model fails if the data are not properly vetted by an operator, maintenance staff, or engineer in the process. It is standard to list the bearings, operating speeds, and power outputs necessary for the system. To leverage more physics-based models on time, it would be prudent to include the system dynamic properties for data verification and analysis.

5.1.4 Bearing Physics Model Response: Idealized

5.1.4.1 Bearing Deformation

The initial model formulation considered the process with the minimum values of parameters from the test system parameter formulation. In addition, the defect modeling assumes that the defect is initially placed at the center of the load zone for the bearing in the inner race defect and as a constant in the outer race defect case. Figure 5.7 shows the rolling element deformation with the inner race defect condition. Each line in the Figure represents a different bearing rolling element. It should be noted that two colors are repeated (yellow and blue) to represent all of the rolling elements. As seen from Figure 5.7, there is an accepted operating deformation limit of the bearing around and between -2E-5 m and 2E-5 m. The band around this deformation range represents the normal bearing vibrations without any defect added. However, with the defect added, an impulsive signal variation symbolizes a rolling element hitting the bearing surface. These rapid signals are evenly spaced to match the bearing inner race characteristic frequency, 151 Hz, for the 6205 bearing

at 1687 RPM.
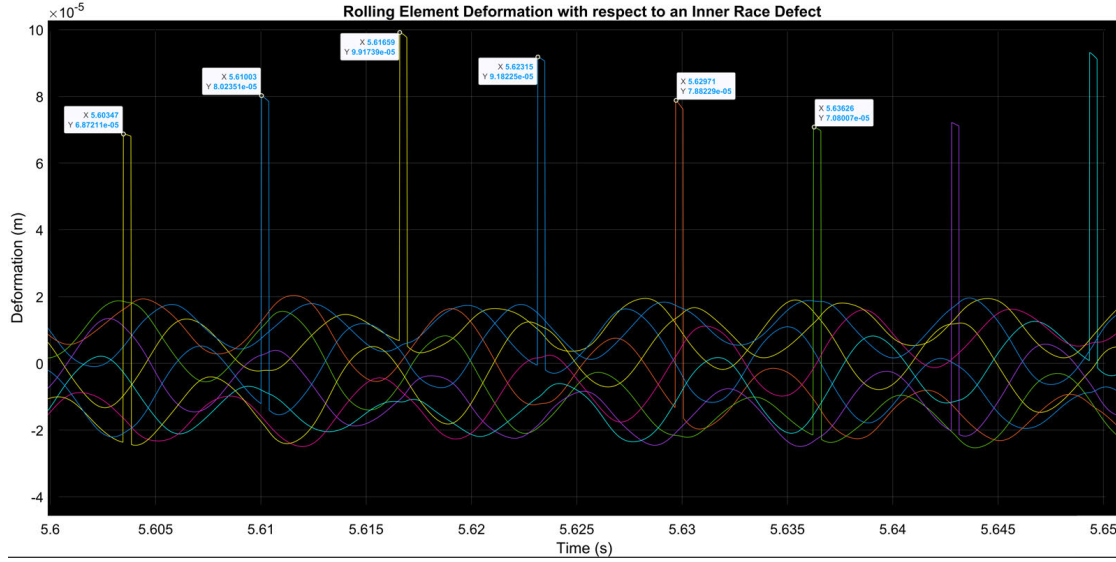
Figure 5.7: 6205 Bearing Rolling Element Deformation - Added Inner Race Defect of 0.1 mm depth and 1mm width

Figure 5.7 shows a different height in the deformation strikes as the rolling elements rotate. These correspond to the positions where the inner race defect may appear as the inner race rotates with the shaft. In addition, it does not occur simultaneously every time because the rolling elements rotate at a slower relative to the inner race. The defect moves at a rate of ω_i, which is the bearing running speed in radians per second (RPS). The ratio is actually dependent on a ratio of the ball diameter (b_D) and pitch diameter (P_D), as shown in Equation 5.33. The equation is actually adapted from the bearing equation FTF, but ω_i is not in Hz.

$$\omega_c = (\frac{\omega_i}{2})(1 - (\frac{b_D}{P_D})) \tag{5.33}$$

As an example, Figure 5.8 represents what would occur if the defect were placed on the outer race. In this instance, the bearing spike occurs in the same position each time the rolling element reaches the top of the curve. The continuous occurrence of the deformation spike in one location corresponds to a defect that does not move with respect to the bearing rotation. Due to the lack of movement, the resulting bearing deformation from the defect is the same with each occurrence. Similar to the inner race defect, the occurrence for each spike matches the expected bearing outer race characteristic frequency (100 Hz).

235
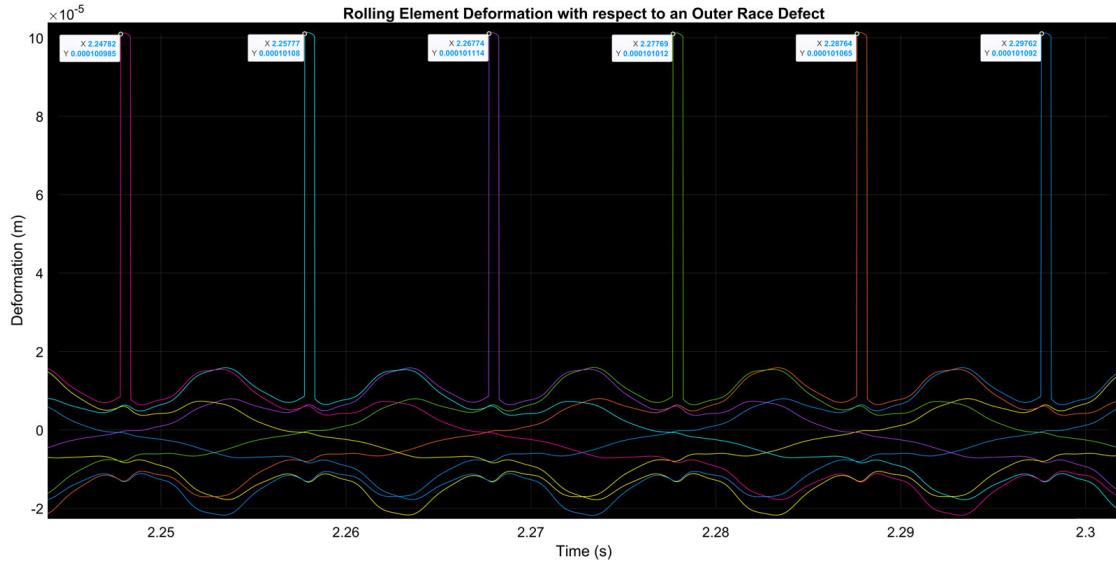
Figure 5.8: 6205 Bearing Rolling Element Deformation - Added Outer Race Defect of 0.1 mm depth and 1mm width

As seen in Figure 5.7 and 5.8, the primary difference between the two different defects is the location of where the rolling element strikes the defect. In the positive region, that represents the rolling element striking the defect in the load zone. In the negative region, the rolling element strikes the defect in the unloaded zone (*i.e.,* where the bearing force is not applied). The resulting difference in the location causes a lower force to act on the bearing according to Equation 5.19, due to the lower resulting deformation. Hence, as a defect worsens over time, the force becomes larger, resulting in a greater defect size.

Figure 5.9 shows the effect on the bearing deformation due to the lengthening defect during the experimental testing. For the 3 mm defect case (Figure 5.9 (b)), the bearing deformation does not appear to increase greater than the original deformation provided in the original case of Figure 5.9 (a). However, from the defect length, it is possible to see that the deformation and the rolling element follow the perceived path of the rolling element based on the resulting curves in the deformation. In addition, the base deformation, the area without the spikes, appears to increase due to the bearing defect lengthening. The most probable cause for the increase appears to stem from the resulting force acting on the bearing from the increased and prolonged deformation period.

For the 7 mm defect range, each rolling element appears to maintain contact with the bearing defect for at least approximately 25% of the rolling element revolution. In the 7 mm defect
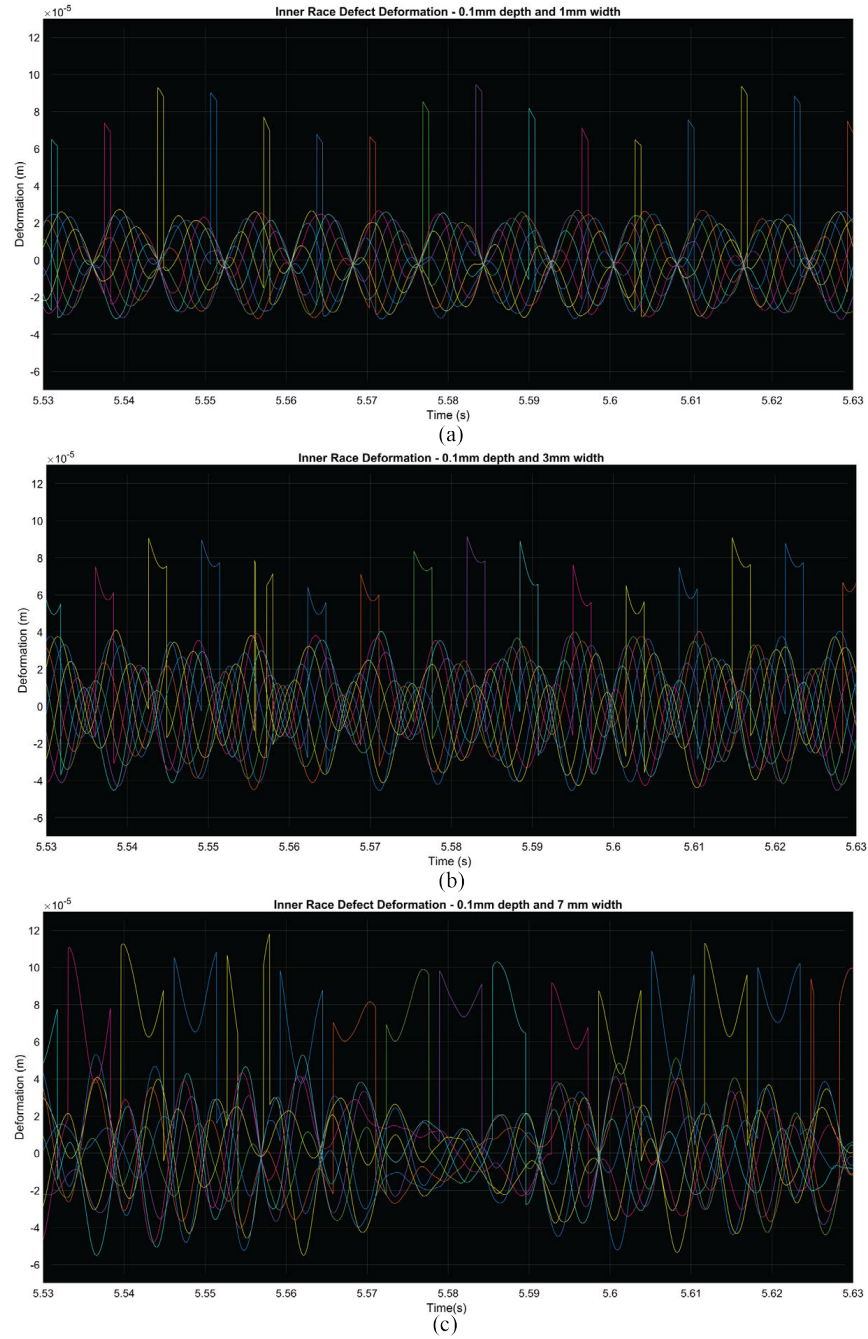
Figure 5.9: Bearing fatigue defect progression effect on deformation (a) 1mm inner race defect width, (b) 3 mm inner race defect width, (c) 7 mm inner race defect width.

case, the base deformation appears to operate between -6E-5 m and 6E-5. That growth in the base deformation is also likely due to the prolonged period that the rolling element is in contact with the defect on the bearing surface. The prolonged period of contact provides a basis for the increase in

237

the bearing defect size that does not necessarily occur with the change in the defect size. However, the deformation increase does appear to slow as the bearing defect grows along the bearing raceway. The deformation periodicity changes due to the lengthening bearing defect shown in Figure 5.9 (c) at around 5.58. The change in deformation could be reflected in the subsequent vibration as a critical point reached in the bearing excitation.

Figure 5.10 contains the deformation values from the bearing rolling elements under different size conditions. In this comparison, the the mass, spring, and damper parameters were kept the same whereas only the bearing geometry and ball stiffness parameters were evaluated and changed. In this instance, the bearing geometry influences the overall bearing deformation in relation to the bearing width and time spent on the deformation. The bearing raceway size kept increasing, whereas the defect width was kept constant. Hence, at the larger bearing sizes, the deformation width is not as large as it was for the smaller bearing sizes. While the variation between, these bearings is 5 mm between the different inner race sizes, consider the case of a 1 mm defect on a 6222 bearing. The inner race diameter is 110 mm. A 1 mm defect is not as severe on the bearing raceway in that instance as it is in the case for a 6205-size bearing.

5.1.4.2 Force Effects

The resulting bearing force is calculated as a summation of the acting deformation on each of the bearing rolling elements. If the deformation is negative, then it is not included in the bearing force approximation based on the Heaviside incorporation for the deformation calculation. However, due to the bearing geometry and the number of rolling elements for the bearings, at least half of the rolling elements are incorporated into the bearing force at any given time. Figure 5.11 shows the bearing deformation and the subsequent effect on the bearing loading for the 1 mm defect width (a) and the 7 mm defect width (b) on the inner race. In the 1mm defect case, there are apparent spikes in the bearing data that correspond to when the rolling element strikes the bearing. Outside of the bearing rolling element striking the bearing defect, the bearing force appears to remain steady about zero.

In the 7 mm defect case (Figure 5.11 (b)), the bearing defect width increases to the scale that at least one rolling element is in contact with the bearing defect at any given time. Hence, the spikes in the bearing load are larger and appear to dominate the bearing force response, in contrast to the 1mm defect case response. In this instance, the bearing force appears to fluctuate
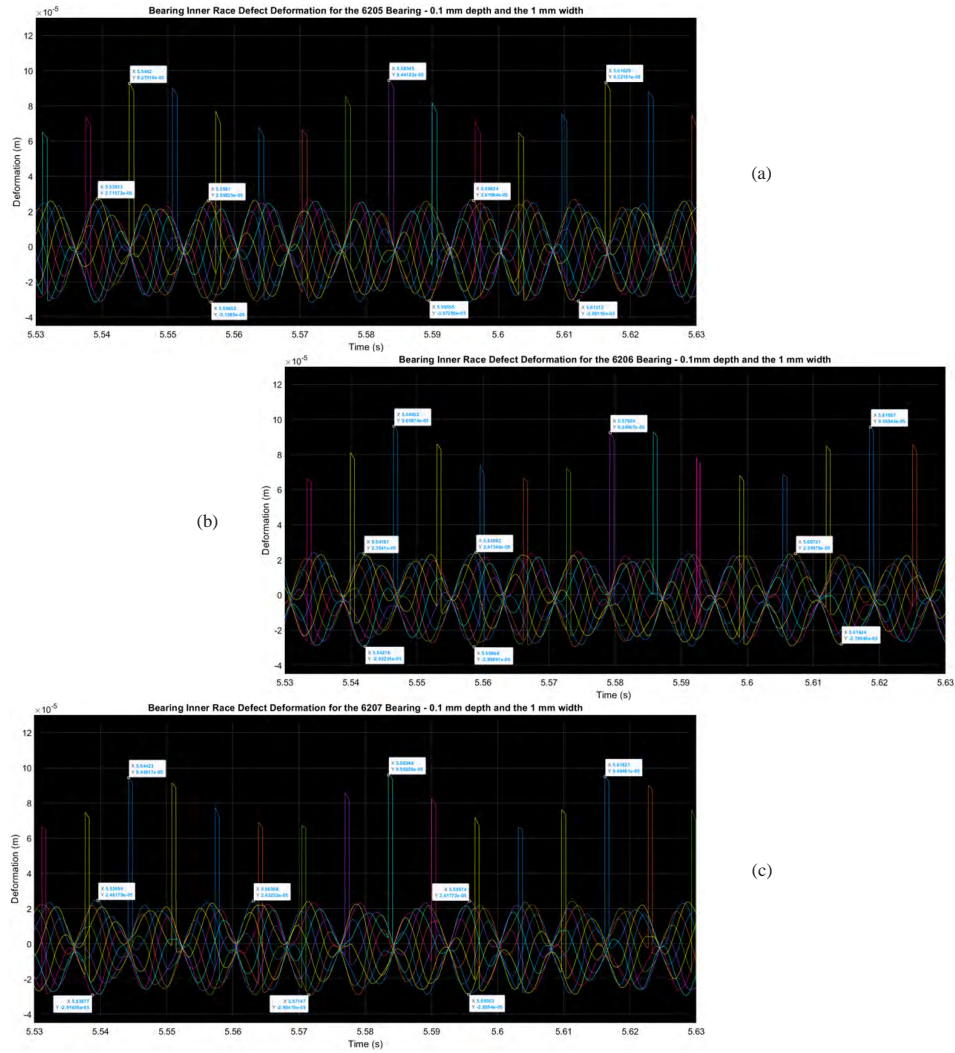
Figure 5.10: Different bearing size deformation under similar physical parameters (a) 6205, (b) 6206, and (c) 6207

between -35 and 35 N in the x-direction and -30 N and 20 N in the y-direction. It should be noted that in both cases, the force sign is expected to change as the bearing elements rotate. In the actual measurement of vibration, the vibrations are expected to oscillate through the system; hence, the force also oscillates. It should be noted that the force is not necessarily centered about zero, maintaining that the bearing does have a constant force acting on it outside of the applied force

239
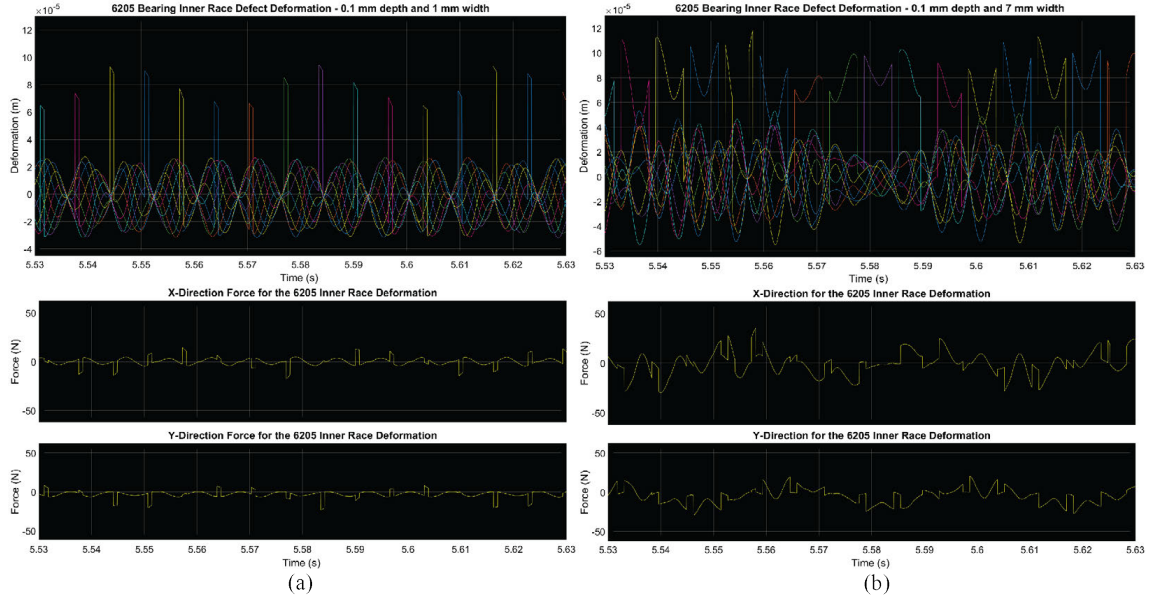
Figure 5.11: 6205 Bearing deformation and resulting load for the inner race (a) 1 mm defect and (b) 7 mm defect width

from the defect impacts.

Again, as a proof of concept and model validation, the outer race defect cases for the 1 mm and 7 mm defects are presented in Figure 5.12. Similar to the outer race deformation observation, the bearing force is consistent as the rolling elements come into contact with the bearing defect. In addition, the x-direction force is much lower than the y-direction force in both cases. Viewing the deformation for both cases, the curves are much smoother than the inner race defect. An observation is that the bearing deformation is consistent because the defect is not moving. The only reason for the increase and change in defect shape is because the defect is growing larger. Even as the defect grows, the deformation remains constant and does not change drastically. The extended time spent at the higher deformation raises the overall average force but does not cause the variation as seen in the inner race defect. In addition, with the defect remaining stationary at the bottom of the raceway, the y-direction forces receive the bulk of the deformation acting on the bearing. As the defect increases in length and begins to creep along the surface of the bearing, the force gradually rises due to the prolonged time spent on the defect.
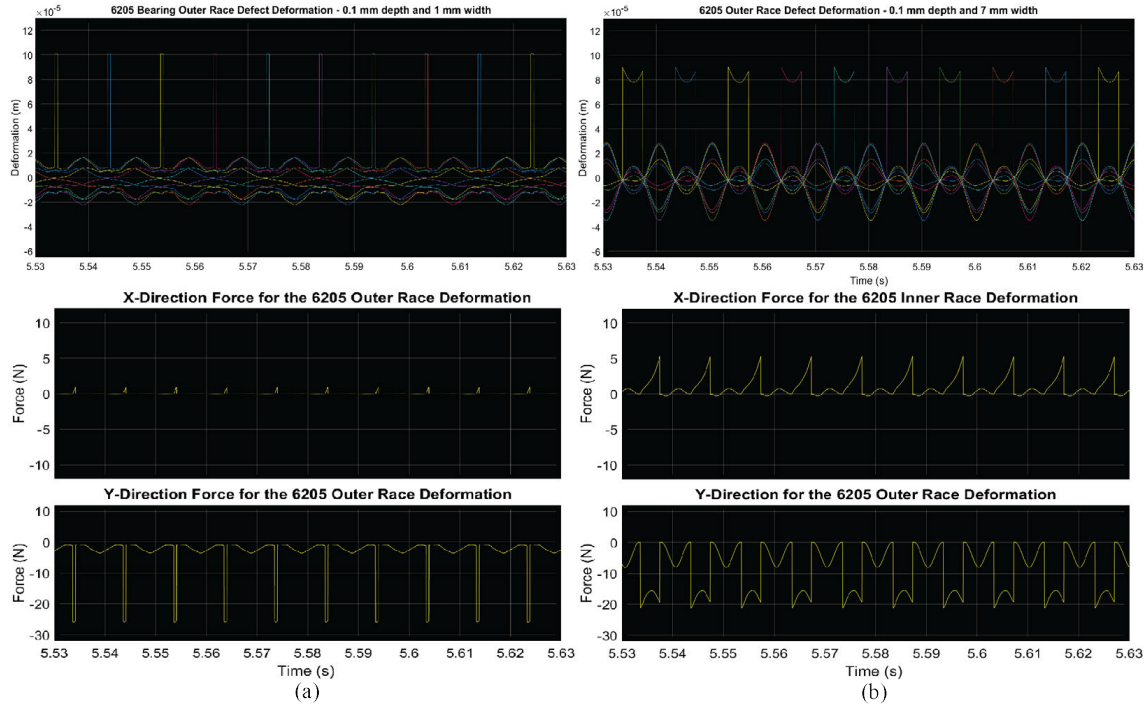
Figure 5.12: 6205 Bearing deformation and resulting load for the outer race (a) 1 mm defect and (b) 7 mm defect width

5.1.4.3   Vibration Effects (Idealized Parameters)

The bearing vibration was assessed relative to the difference of the bearing pedestal and shaft in both the x- and y-directions. The difference in these both system elements serves as the basis for the deformation approximation for the bearing. As mentioned in Section E.1, the parameter configuration varies based on the method chosen to represent the rolling elements. Furthermore, a contributing factor is the number of unknowns observed in the experimental data, such as temperature, measured grease viscosity, preload and bearing clearance. Additionally, a further degree of experimental noise could have stemmed from the test stand assembly. If any of the experimental system became loose during testing and were not captured during testing. Hence, the vibration effects from the physics models represent the "ideal" approximation of the parameters tested based on hte physics parameter selection. In this instance, a broad spectrum of parameters are tested.

The general rule for testing for the parameters did follow these assumptions:

i.) The mass, bearing geometry, and ball stiffness was kept constant despite possible changes to the system. These were shown to have little change based on the overall system.

241

ii.) The pedestal and shaft stiffness are constant throughout there respective bodies and do not change relative to the x- or y- direction. The tested range for the pedestal stiffness ranges from 1E7 $\frac{N}{m}$ from Mishra et $al.$ [516] for a similar bearing size and to the maximum possible material stiffness 2E11 $\frac{N}{m^2}$. The shaft stiffness is tested in the range 3E5 based on an Ansys output for the minimum value and then up to a possible maximum 1e8 $\frac{N}{m}$ from Mishra et $al.$ [516].

iii.) The damping parameters are configured based on Equation 5.30 from Genta et $al.$ [519] and Crandall et $al.$ [518]. The loss factor for the material is set at 0.01 for the metal structures.

iv.) The resonance parameters are not changed in the initial testing and formulation.

The simulations were run in MATLAB Simulink and formulated based on the equation set from Equation 5.15. The equations of motion were solved using the ODE45 solver. The ODE 45 solver uses the Runge-Kutta formulation to solve the equations of motion. The problem and solver begins by reducing the problem to the format of a series of first-order equations. A condition was included if the inner or outer race defects were used in the model analysis. After conducting model tuning, the final parameters were selected based on the model response to the baseline configuration of the test stand. In this essence, though, the model was tuned to as close an approximation as possible and is still subjected to system error. Table 5.9 considers the system elements from the bearing system with the same damping ratios across the different bearing sizes for the pedestal (ξ_p = 0.9) and shaft (ξ_s = 9). Table 5.4 shows the stiffness parameters used. It should be noted that an additional scaling factor of 1000 was used to limit results and prevent overshoot in the back end calculation of the force. It is not certain what was causing the overshoot as it continued to occur even after scaling the parameters.

Table 5.9: CGEC Test System: Final Model Parameters

Parameters	6205	6206	6207
K_p	4.920e10 $\frac{N}{m}$	6.954e10 $\frac{N}{m}$	6.954e10 $\frac{N}{m}$
K_s	4.442e6 $\frac{N}{m}$	4.398e6 $\frac{N}{m}$	4.122e6 $\frac{N}{m}$
R_p	4.438e3 $\frac{Ns}{m}$	5.266e3 $\frac{Ns}{m}$	5.266e3 $\frac{Ns}{m}$
R_s	266.172 $\frac{Ns}{m}$	289.321 $\frac{Ns}{m}$	292.8857 $\frac{Ns}{m}$

Figure 5.13 contains the theoretical bearing response for the bearing response in both the x-direction and y-direction for the tested bearing sizes. The bearing vibration is measured based on a point on the inner race defect. The impulse events placed on the bearing are based on the

elastic deformation from the rolling element passing over the bearing surface on the inner race. The difference in the x- and y-direction vibrations stems from the impulse events on the bearing surface. It appears, based on the model appearance, that the y-direction vibration is more dependent on the bearing response, whereas the x-direction vibration is affected by changes to the system parameters. Interestingly, only the x-direction vibrations appear between the different cases, which could correspond to the difference in the system parameters. The system parameters in the 6206 and 6207 cases appear very similar, whereas the difference in the damping elements for the 6205 case could cause variation.
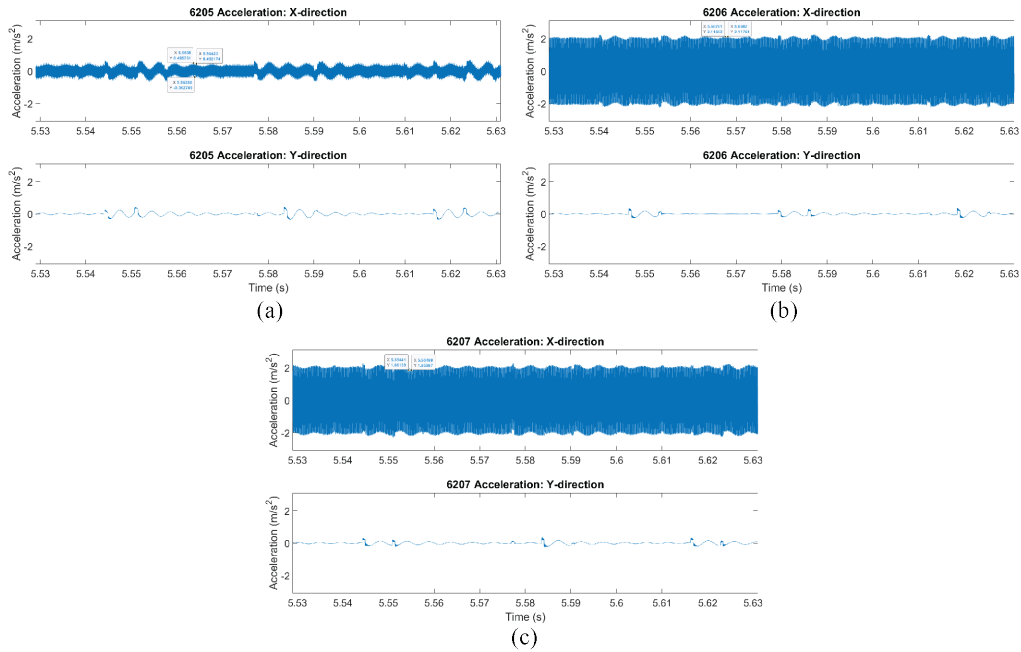
Figure 5.13: Output of the baseline vibration for the different bearing sizes: (a) 6205, (b) 6206, and (c) 6207

Figure 5.14 shows the 6205 bearing vibration when the vibration output is similar to the 6206 and 6207 bearing vibration. The stiffness values remain the same, but the damping ratios are changed to 0.1 for the pedestal and 15 for the shaft, respectively. The change to the damping ratios causes the pedestal damping value (R_p) to be 493.1 $\frac{Ns}{m}$ and the shaft damping value (R_s) to 443.6 $\frac{Ns}{m}$. The likely change to the system vibration could stem from changes to the system's natural and damping frequencies, which in turn cause the change to the x-direction vibration. Incidentally, it should still be noted the y-direction velocity does not yet change, signifying a lack of effect from the system impulses. Hence, it is possible to consider that the defect effects of the bearing may seem

243

similar in the case of the y-direction despite the changes in stiffness and the bearing deformation between different bearing sizes.
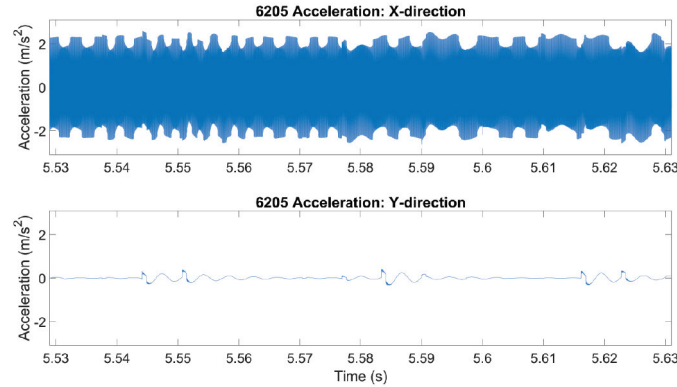
Figure 5.14: Output of the baseline vibration for the 6205 bearing at a different damping ratio for the shaft and pedestal

Figure 5.15 compares the experimental and simulated bearing data over the same period as the deformation, force, and earlier vibration investigation. The simulated bearing data does not appear to vary as the experimental system would. In the horizontal experimental data, the collected bearing data fluctuates around the 0 value, suggesting that several random impulses are injected into the signal that prevent a steady reading, such as the block not remaining steady or the loosening of the sensor over time. These experimental observations could reflect unanticipated noise in the system. In the y-direction simulated data from Figure 5.13 and 5.14, it would appear that there can be some anticipated fluctuation around 0.

From Figure 5.15, Figure 5.15 (a) and Figure 5.15 (b) shows the 6205 experimental vibrations compared to the simulated vibration data. Based on the variation in the data, the expected bearing output appears to match closer to Figure 5.15 (b) than the output from Figure 5.15 (a). The random impulse could signify unexpected noise from the data in this scenario. For Figure 5.15 (c) and Figure 5.15 (d), the heightened impulse events appear to exceed the simulated bearing vibration limits. These, in turn, could be filtered out based on the simulation output.

Figure 5.16 shows the raw simulation output for the different bearing defects. Based on the vibration output, the defect progression seems intermixed; however, as the defect progresses, the maximum possible amplitude does increase. The cause of the vibration increase stems from the larger deformation period that the bearing experiences. The impulses are distinct in the initial bearing
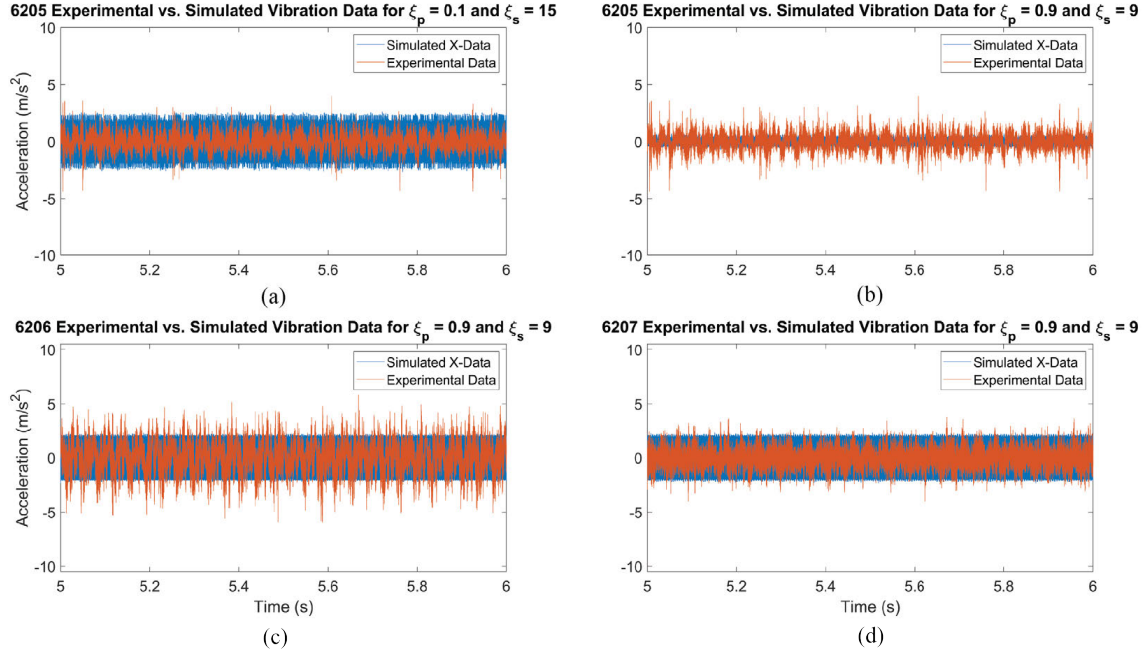
Figure 5.15: Bearing simulated versus experimental comparison: (a) 6205 experimental vs. simulation vibration based on damping pedestal ($\xi_p = 0.1$) and shaft ($\xi_s = 15$), (b) 6205 experimental vs. simulation vibration based on damping pedestal ($\xi_p = 0.9$) and shaft ($\xi_s = 9$), (c) 6206 experimental vs. simulation vibration based on damping pedestal ($\xi_p = 0.9$) and shaft ($\xi_s = 9$), (d) 6207 experimental vs. simulation vibration based on damping pedestal ($\xi_p = 0.9$) and shaft ($\xi_s = 9$)

case and occur at the bearing characteristic frequency for the inner race defect case 151 Hz. The impulses should then match the expected response noted from the investigation of the deformation effects and modeling shown in Section 5.1.4.1. Table 5.10 shows the statistical characteristics of the bearing data. As the bearing defect increases, the initial case (1mm) is much greater than the baseline case by 5 m/s^2. After the initial increase, the increase appears much lower as the bearing defect progresses from the 1 mm to the 3 mm case and then to the 7 mm case.

A similar trend occurs in the data variance. The heightened response indicates a greater dispersion of values relative to the peak, meaning a possible increase in the outliers. The increase in outliers would match the expected increased response in the bearing impulse due to a greater defect. Interestingly, the kurtosis value does not necessarily experience a change relative to the increased defect size, except in the 6205 case. The decrease could indicate a trend; however, in the other bearing size, the kurtosis values were not skewed in a similar way. The lack of change could indicate that the size of the parameters may influence the vibration results.

Figure 5.17 shows the simulated response when the bearing data are downsampled to 40
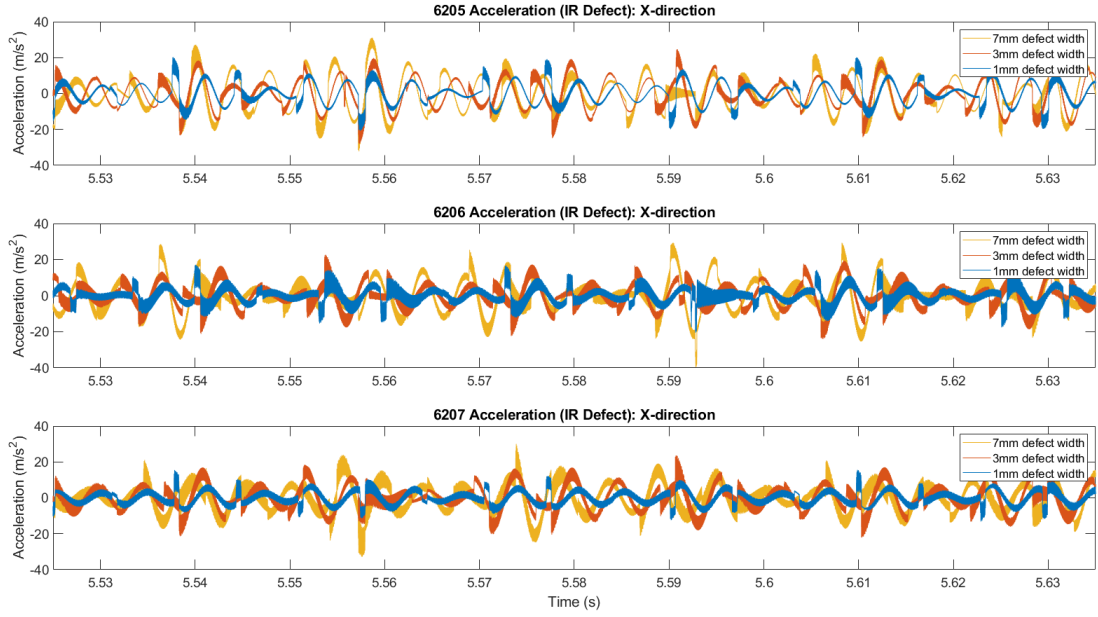
Figure 5.16: Simulated bearing vibration for the different defects stages on all tested bearing sizes

Table 5.10: Bearing Raw Vibration Statistical Characteristics

Bearing Condition	RMS	Variance	Kurtosis	Skewness
6205 I.R. 1 mm Defect	6.3610 $\frac{m}{s^2}$	40.4630 $(\frac{m}{s^2})^2$	3.0031	-0.0046
6205 I.R. 3 mm Defect	8.4823 $\frac{m}{s^2}$	71.9494 $(\frac{m}{s^2})^2$	2.5581	-0.0288
6205 I.R. 7 mm Defect	10.862 $\frac{m}{s^2}$	114.117 $(\frac{m}{s^2})^2$	2.4921	0.0620
6206 I.R. 1 mm Defect	4.6816 $\frac{m}{s^2}$	21.9178 $(\frac{m}{s^2})^2$	3.8753	0.0131
6206 I.R. 3 mm Defect	7.0151 $\frac{m}{s^2}$	49.2115 $(\frac{m}{s^2})^2$	3.2296	-0.0222
6206 I.R. 7 mm Defect	8.7989 $\frac{m}{s^2}$	77.4196 $(\frac{m}{s^2})^2$	3.3424	-0.0095
6207 I.R. 1 mm Defect	4.0307 $\frac{m}{s^2}$	16.2466 $(\frac{m}{s^2})^2$	2.9783	0.0819
6207 I.R. 3 mm Defect	7.2178 $\frac{m}{s^2}$	52.0968 $(\frac{m}{s^2})^2$	2.8861	0.0111
6207 I.R. 7 mm Defect	8.4438 $\frac{m}{s^2}$	77.2974 $(\frac{m}{s^2})^2$	2.9365	-0.0710

kS/s. The resampled simulation bearing response matches the same curve as the raw data. The primary reason for resampling the data is to standardize the sampling rate throughout the signal. When using the ODE 45 solver, the step size varies, indicating that the time step differs between the samples. The data resampling is to try and create a standardized step rate so that it is possible to

use frequency techniques. Based on the resampled data, the trend appears similar to the response to the bearing baseline data. However, the only difference now that causes the changes in amplitude may stem from the bearing defect cases.
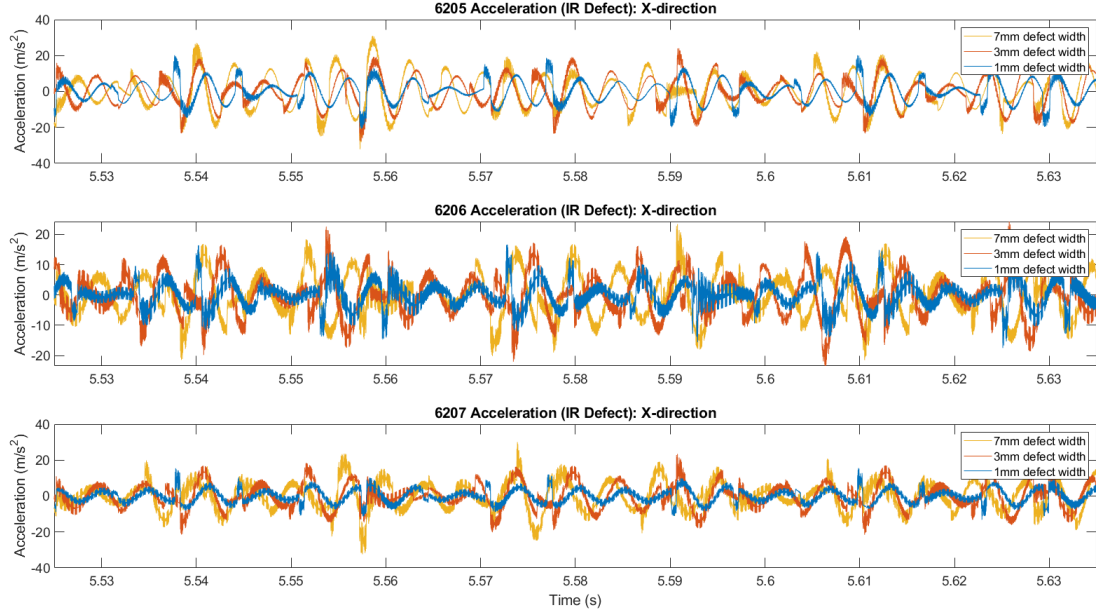
Figure 5.17: Simulated bearing vibration for the different defects stages on all tested bearing sizes at a 40 kHz sampling rate

However, there is some relative error in performing the resampling. In this instance, a separate time vector was created based on the desired sampling rate. The selected vibration value was based on the closest value to each time value in the desired vector. Figure 5.18 shows the difference between the closest time values from the simulation and the new time values based on the sampling rate. The time difference between samples is 2.5E-5 s for a 40 kS/s rate. The maximum difference found across the different bearing cases was approximately 8E-6 for each bearing case. In that instance, the replaced point could be up to 8 microseconds from the expected position. On average, though, the error is approximately 3E-6 for the different bearing cases. To compensate for the possible error, a margin of error is needed based on the difference in the sampling rate. The most likely margin of error is to match up the potential defects for the bearing response based on the FFT peaks. The peaks can be scaled based on the expected impacts. However, the location error should be noted based on an expected impact, such as the running speed and the bearing defect characteristic frequencies. In addition, due to the data drift in the y-axis, the data also need to be

247

centered about 0 for the FFT response.
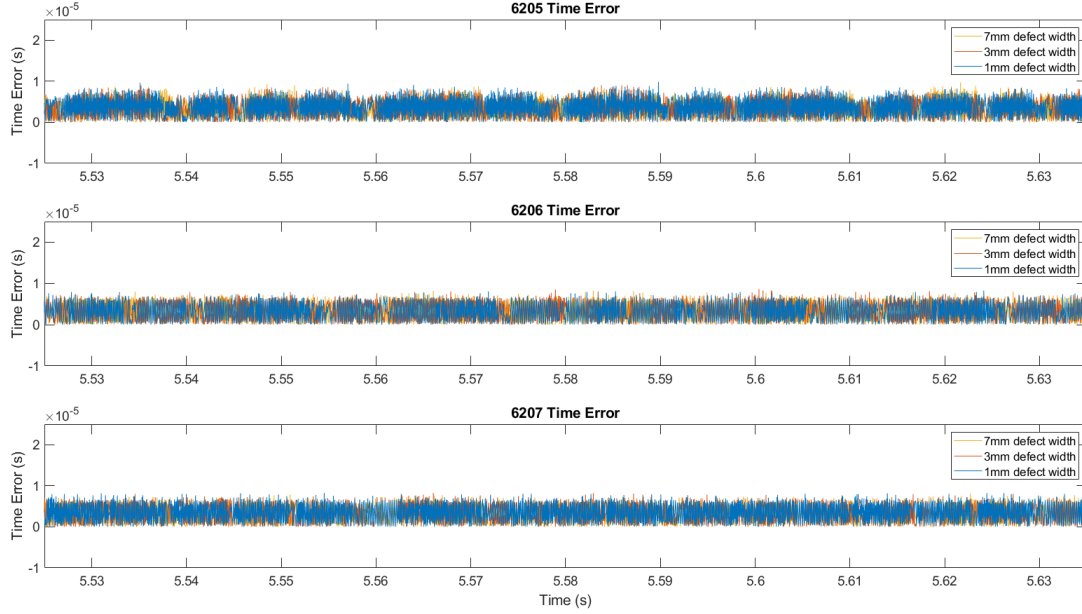
Figure 5.18: Measured the difference between the expected and simulated time value

Table 5.11 contains the statistical characteristics for the different simulated bearing responses for the resampled bearing data. Overall, there was a general decrease in the RMS values across the different bearing stages. The 6206 bearing data showed a very large decrease in the 7 mm defect case and the variance. The decrease could have resulted from the bearing simulated response since the 6205 and 6207 data are only marginally affected concerning that bearing system. The bearing kurtosis does not appear to change from the decreasing trend except for in the 6207 case. The data skewness does not appear affected by the change in the amount of vibration simulated data.

The raw and resampled vibration data were adjusted for the average at 0. However, the mean difference in the data was only approximately 0.01 at its greatest point across the different simulations. Figure 5.19 shows the FFT with the unsampled simulation bearing response with the frequency response scaled based on the expected impact response rather than using the raw FFT value. It should be noted that the original response was initially in the range of 400 Hz, with little frequency response in the lower frequency response. The overall error in the sampling response may cause this. In an idealized scenario, more of a spike was expected in the time series response, whereas here, the approximation appears to allow for a general reduction. The point of the scaling

248

Table 5.11: Bearing Resampled Vibration Statistical Characteristics

Bearing Condition	RMS	Variance	Kurtosis	Skewness
6205 I.R. 1 mm Defect	6.0593 $\frac{m}{s^2}$	37.1526 $(\frac{m}{s^2})^2$	3.0374	-0.0026
6205 I.R. 3 mm Defect	8.3428 $\frac{m}{s^2}$	69.6022 $(\frac{m}{s^2})^2$	2.5379	-0.0290
6205 I.R. 7 mm Defect	10.373 $\frac{m}{s^2}$	107.605 $(\frac{m}{s^2})^2$	2.4691	0.0455
6206 I.R. 1 mm Defect	4.5863 $\frac{m}{s^2}$	21.034 $(\frac{m}{s^2})^2$	3.8850	0.0091
6206 I.R. 3 mm Defect	6.9310 $\frac{m}{s^2}$	48.0393 $(\frac{m}{s^2})^2$	3.2374	-0.0219
6206 I.R. 7 mm Defect	7.1497 $\frac{m}{s^2}$	51.1185 $(\frac{m}{s^2})^2$	2.8819	0.0134
6207 I.R. 1 mm Defect	3.9954 $\frac{m}{s^2}$	15.9530 $(\frac{m}{s^2})^2$	2.9582	0.0786
6207 I.R. 3 mm Defect	7.1497 $\frac{m}{s^2}$	51.1185 $(\frac{m}{s^2})^2$	2.8819	0.0134
6207 I.R. 7 mm Defect	8.3724 $\frac{m}{s^2}$	70.0974 $(\frac{m}{s^2})^2$	2.9182	-0.0700

was to see that, despite the error the model amplitude was in the general range of the sensor values from the experimental data. The scaling of the different values allowed for a better one-to-one comparison to the experimental data. Figure 5.20 shows the FFT with the resampled simulation bearing response with the frequency response scaled based on the impact response. In all cases, the frequency data needed to be scaled to the frequency impact point. The maximum error within the frequency response in terms of frequency impact varied from 6% to 80%.

The most error came from the resampled data. Hence, the error of the bearing resampled may affect the overall impact of the bearing data response. For the FFT to even work for the raw vibration output, the bearing data were divided based on timestamp. The sampling rate was based on averaging the data to the nearest two significant figures. Depending on the simulated response, the bearing frequency amplitude was within the range 68 kHz–77 kHz. There was never more than an error of 400 samples per second, which was less than a 0.6% error. The extra samples were considered negligible for the overall analysis.

Despite that frequency shift in the Fourier analysis, the other noticeable change in the bearing data between the raw vibration output and the resampled vibration data was the number of additional frequency peaks in the FFTs. Based on the difference in the vibration output for the raw and resampled simulation data, the additional peaks may be residual effects of the inner race defect as it tries to reach the expected smoother output. Considering Figure 5.16 and Figure
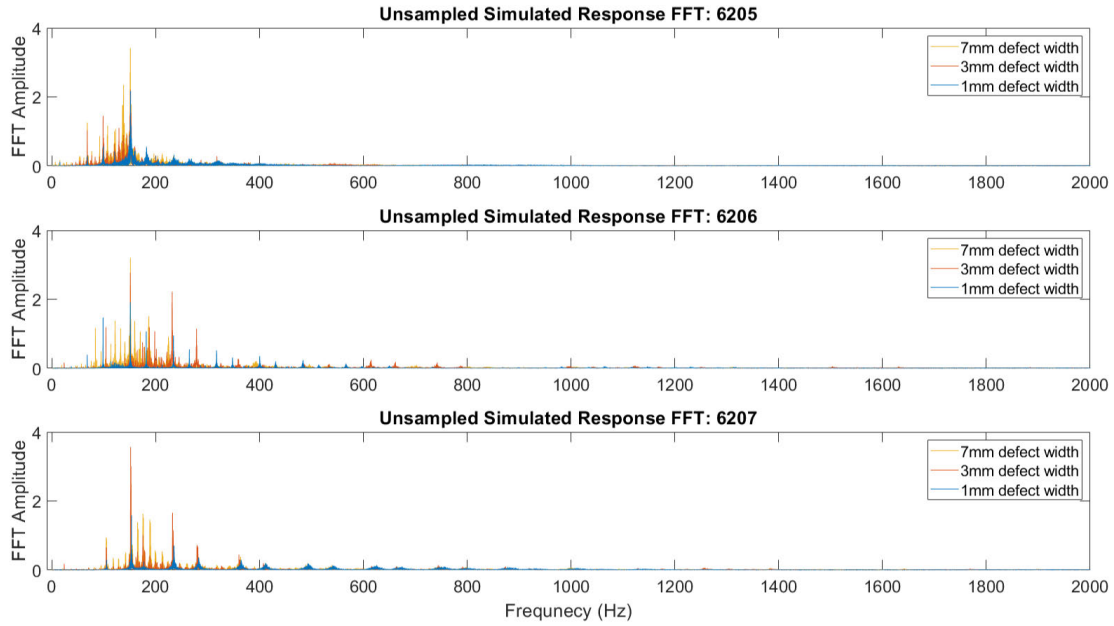
Figure 5.19: FFT from the unsampled vibration response data for the different bearing cases at 1687 RPM

5.17, which are the time series representations for the simulated bearing data, approximately 30% to 40%, depending on the simulation response, of the total raw data are removed. Hence, the removed chop in the data could affect the expected bearing output for the system despite standardizing the sampling rate.

Based on the frequency impacts, the other item noted is that the simulation bearing response is much greater than the expected vibration response for the different defect cases as observed experimentally. As noted in the average fatigue FFT from Figure 4.75, the maximum amplitude is just above 0.6 $\frac{m}{s^2}$ from the frequency amplitude. In the simulation response, the maximum amplitude recorded is 4 $\frac{m}{s^2}$. In addition, there was no response at higher frequencies greater than 500, or they were much lower. A possible cause for this is a different frequency response in the model than for the collected bearing data.

Figure 5.21 compares the defect cases at a different frequency response (1000 Hz) for the experimental data compared to the simulated response from the raw data. The high amplitude shown at 60 Hz for the experimental data did lead to the consideration that the simulated data peak is closer to the 60 Hz location than the defect frequency cases. However, by shifting the data to the 60 Hz point, there is no frequency response in the defect range of the bearing, which is not
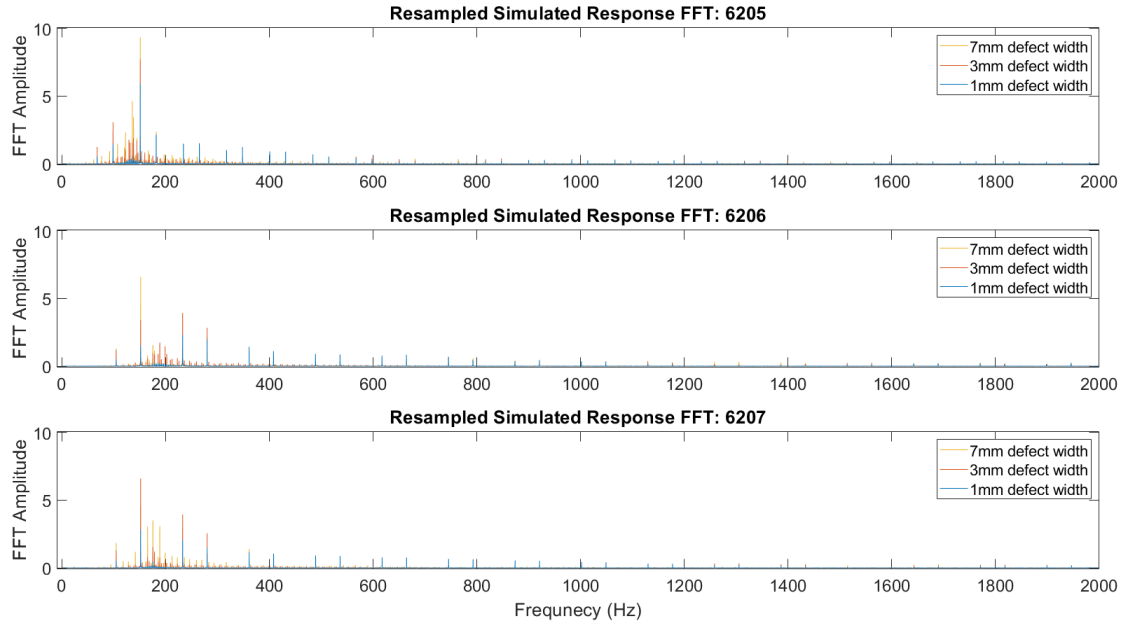
Figure 5.20: FFT from the resampled vibration response data for the different bearing cases at 1687 RPM

plausible given the system. Hence, despite the variation in the peak amplitude around the bearing defect frequency, the frequency response change appears to show closer corroboration to the expected bearing response after the scaling. However, that is not necessarily valid in an actual application due to the infeasibility of applying such a scaling to raw data.
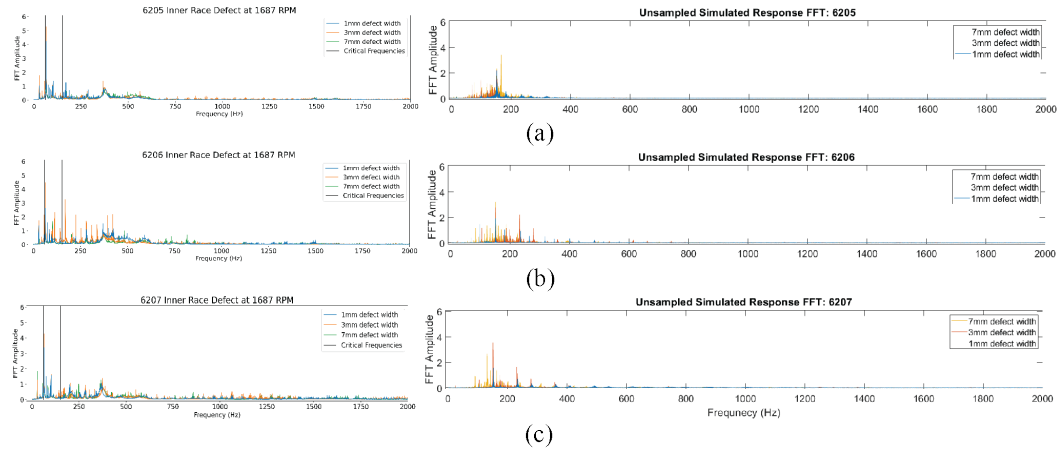
Figure 5.21: Comparison of the inner race defect case from the experimental (left) and simulated response (right) for the different bearing sizes: (a) 6205, (b) 6206, and (c) 6207

In each case, the peak error was evaluated at the specific frequency indices with the observed

experimental data and the simulated MATLAB data. Table 5.12 contains the percent error between the frequency peaks at 151 Hz for the simulated and experimental data and the percent error between the RMS values in the experimental and simulated values. The high error expected between the simulated and experimental data is not unexpected, mainly due to the frequency scaling. There is not necessarily enough to determine another method to scale the frequency data appropriately otherwise.

Table 5.12: Error Percentage based on frequency peaks and RMS difference between the experimental data and raw vibration data without frequency shift

Error Category	IR 1 mm	IR 3 mm	IR 7 mm
6205 Freq. Peaks (FS)	88.21%	88.21%	86.42%
6206 Freq. Peaks (FS)	91.42%	52.50%	94.22%
6207 Freq. Peaks (FS)	85.68%	83.34%	92.88%
6205 RMS Difference	7.10%	25.05%	81.98%
6206 RMS Difference	35.62%	21.50%	61.78%
6207 RMS Difference	33.64%	8.68%	28.60%

Hence, the frequency peak comparison does not determine whether there is an accurate representation between the model and experimental responses. The RMS error noted over the different time series yielded a better error approximation on average. However, there were notable outliers in the IR 7 mm case for the 6205 and 6206 data. However, a more similar RMS implies that, despite the possible peak outliers, on average, the model response shows similarities for select bearing responses, such as the 6205 IR 1mm and 5207 IR 3mm case. A statistical test (Kruskal Wallis) was also considered to determine if there was similarity between the experimental and simulated data. The Kruskal-Wallis test is a one-way ANOVA (Analysis of Variance) test that is a non-parametric method for testing whether the samples originate from the same distribution. The null hypothesis is that the different sample groups provided to the test are from the same distribution. The test is chosen because it can handle equal or different sample sizes. A random data subset is taken from the experimental data to match the same size as the simulation data. However, none of the experimental data appeared to suggest that they came from the same distribution as the simulation data.

Due to the variation in the time and frequency content, the model and data did not relate under these ideal parameters. Hence, an additional round of parameter tuning was conducted based on the experimental results observed from Chapter 4. The overall purpose of the section is to provide

an understanding of how to configure the initial parameters for the chosen bearing model. Future work could consider further ranges for parameters and their relationships to ensure the expected outcome in comparison regardless of sensor type.

5.1.5 Bearing Physics Scaling Evaluation: Experimental Based

With the creation of the physical models for each bearing size and an understanding of the expected values, a decision needs to be made on implementing the physics scaling method. In this application, a high degree of error was noted in the physical model versus the experimental results when considering the collected data from the low-frequency cutoff sensors and even more so when compared to the vibration values from the data generation section. Despite the scaling and moving the frequency representation based on expected frequency phenomena, it did not yield the expected results. It is not a valid consideration when considering a real-world system where even less may be known. In addition, the frequency transform did not match the expected research literature results. Hence, the physical model parameters were further tuned to the experimental results of the original horizontal vibration data to determine what the expected parameters may be and their underlying comparison.

5.1.6 Taguchi Study: Damping Values

Based on the bearing analysis, the parameters that most likely change case-to-case are the bearing damping values. Damping is generally modeled by incorporating a damping ratio or constants from the mass and stiffness values to form a damping matrix. Hence, the damping values were varied from case to case to determine how they affected the model results. It should be noted that the stiffness values recorded earlier in Section 5.1.4 are much higher than what is expected in prior literature. However, these values are considered fixed to allow for tuning these damping parameters, which have more variability given the interactions between the different system properties.

The bearing data were compared to the vertical physical data in the prior model response. However, the vertical frequency data does not change concerning the change in speed. Hence, these data are discarded and compared to the vertical frequency data, which had a higher frequency cutoff for the sensor. The parameters are approximated using a Taguchi analysis to test the different bearing cases. The Taguchi method involves creating an orthogonal array and testing the model

under different parameters to experiment and extract different model responses[523]. It is then possible to correlate certain model outputs to parameter changes and tune the model output to the experimental output.

In this vein, damping parameters were determined by considering a 10-level and two-variable orthogonal array. The number of levels was configured to reduce the computational strain and to adjust search parameters based on the model results within that array. The orthogonal array created a total of 100 combinations to test. The other parameters were kept consistent. The model output was 10 seconds, as the model typically reached a steady state after 5 seconds in early testing. The sampling time was doubled to ensure that the trend continued. The comparison occurred based on the RMS values to the horizontal bearing data as collected in Chapter 4.3 for the failure comparisons for the Taguchi analysis. After the Taguchi analysis constrained the search region, Manual tuning was conducted until the parameters reached a similar value within 0.01 of the RMS values or 0.01 of the variance values.

Figure 5.22 provides an example of the results gathered from the Taguchi array. The results from this array explored a very small search grid under expected damping parameters. Each point represents a test point for the pedestal and shaft damping parameters. The underlying points show examples of key outliers in the data to help constrain or expand that search grid. From the parameters shown in Figure 5.22, the search parameter was expanded to encompass much greater values as it was shown that the bulk of parameters tested in this range did not appear to conform to the results.

Table 5.13 contains the different damping parameters to determine a similar output to the RMS and Variance values. The belief was that a similar response in either would retain a similar data shape. Hence, the process after the initial Taguchi analysis to locate a manual tunable range involved modifying the spring damping variable first to either within 0.1 of the expected RMS or variance value. The change in the pedestal damping parameter was then tuned to the 0.01 of the expected value. The damping parameters match the expected criteria from the tuning process. However, it should be noted that these damping parameters are substantially larger than the expected values from the vertical testing on the order of 100 to 500 in some instances. Interestingly as well, there is no discernible trend in the damping parameters that was expected. The change in the 6206 and 6207 data appears more related to the change in the shaft parameters and could stem from the difference in the shaft weight that was experienced either; although, the difference is very slight.
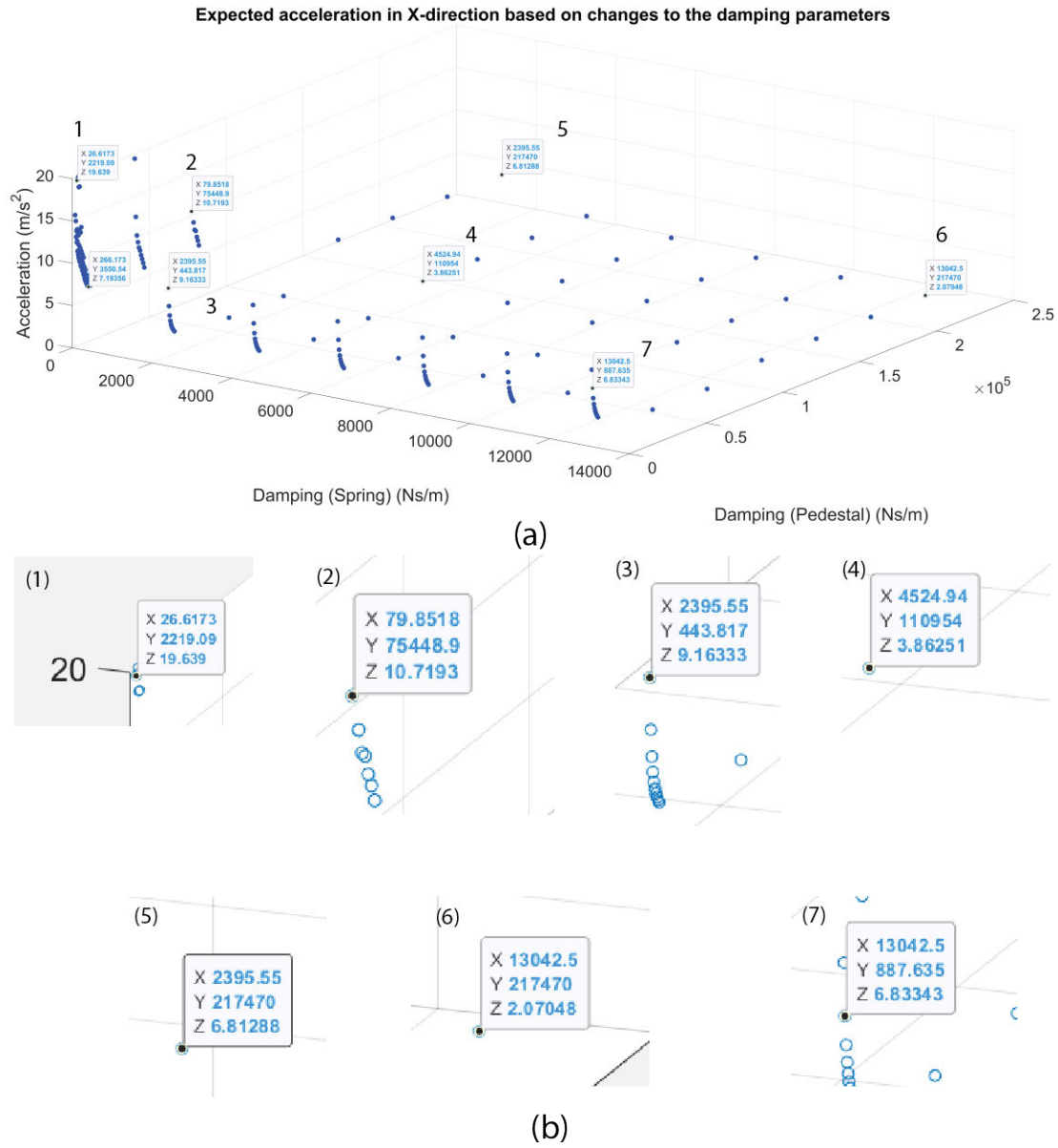
Figure 5.22: Change in RMS X values as the damping parameters are varied for the 6205 parameters. (a) shows the full model run with all of the different damping values under $1000 \frac{m}{s^2}$ and (b) designated points from the simulation of interest

Figure 5.23 shows the sensitivity of the damping parameters by providing an example from the 6205 case with a 7 mm I.R. defect on the raceway. The pedestal damping had a much larger value associated with it, and when it changes from close to 0 to a much greater value, there is a larger change. However, as seen in the figure, the value changes are much smaller when compared to the spring damping parameters. The observation appears to align with expectations as the fixed

Table 5.13: Damping parameters to tune raw experimental data to RMS and variance values

Damping Values	IR 1 mm	IR 3 mm	IR 7 mm
6205 R_p	4.93e5 $\frac{Ns}{m}$	7.93e5 $\frac{Ns}{m}$	7.93e5 $\frac{Ns}{m}$
6206 R_p	3.77e6 $\frac{Ns}{m}$	5.39e5 $\frac{Ns}{m}$	5.39e5 $\frac{Ns}{m}$
6207 R_p	5.93e5 $\frac{Ns}{m}$	5.93e5 $\frac{Ns}{m}$	5.93e5 $\frac{Ns}{m}$
6205 R_s	1.77e4 $\frac{Ns}{m}$	1.45e4 $\frac{Ns}{m}$	1.71e4 $\frac{Ns}{m}$
6206 R_s	2.89e4 $\frac{Ns}{m}$	3.89e3 $\frac{Ns}{m}$	1.15e4 $\frac{Ns}{m}$
6207 R_s	4.23e3 $\frac{Ns}{m}$	1.29e4 $\frac{Ns}{m}$	1.22e4 $\frac{Ns}{m}$

base associated with most pedestals would remain relatively high. Hence, the majority of shaft vibration would appear to occur around the shaft due to the greater amount of flexibility and, have a lower damping value. The same range of weights applied to the damping value changed the value significantly; however, there was a greater change in the acceleration due to the changing weights.

Due to the variability in their assignment, it should be noted that other potential damping parameters are possible within the search space, given the change in the weighting values. The parameters above represent the first parameters found within the search. As shown with the sensitivity plot, certain values can increase or decrease the sensitivity of the results comparatively. Damping, in itself, is a fuzzy parameter as it does rely on system specific aspects, such as the rigidity of the structure, the lubrication in the bearing, and the environmental conditions. A general assumption can be made based on idealized results, but further testing or tuning is needed to ensure an accurate representation. Generally, incorporating a damping ratio accounts for the possible system variation. This tuning accounts for that representation.

5.1.7 Model Prediction for Vibration Response

The results of forcing the damping parameters of the model to yield an output similar to the time series response yielded a better frequency result at the expense of the time series results. Figure 5.24 contains the time and frequency response for the different inner race defect cases for the 6205 bearing size for maintaining a similar variance and RMS. From the different plots, the time plot spikes are much more distinct as compared to the time series plots from the prior section (Section 5.1.4). The underlying noise from the original model response is much more reduced in comparison as compared to the time series data that now matches the time series raw data from the experimental system.

Despite the apparent changes in the time-series plot, the frequency plots appear to conform
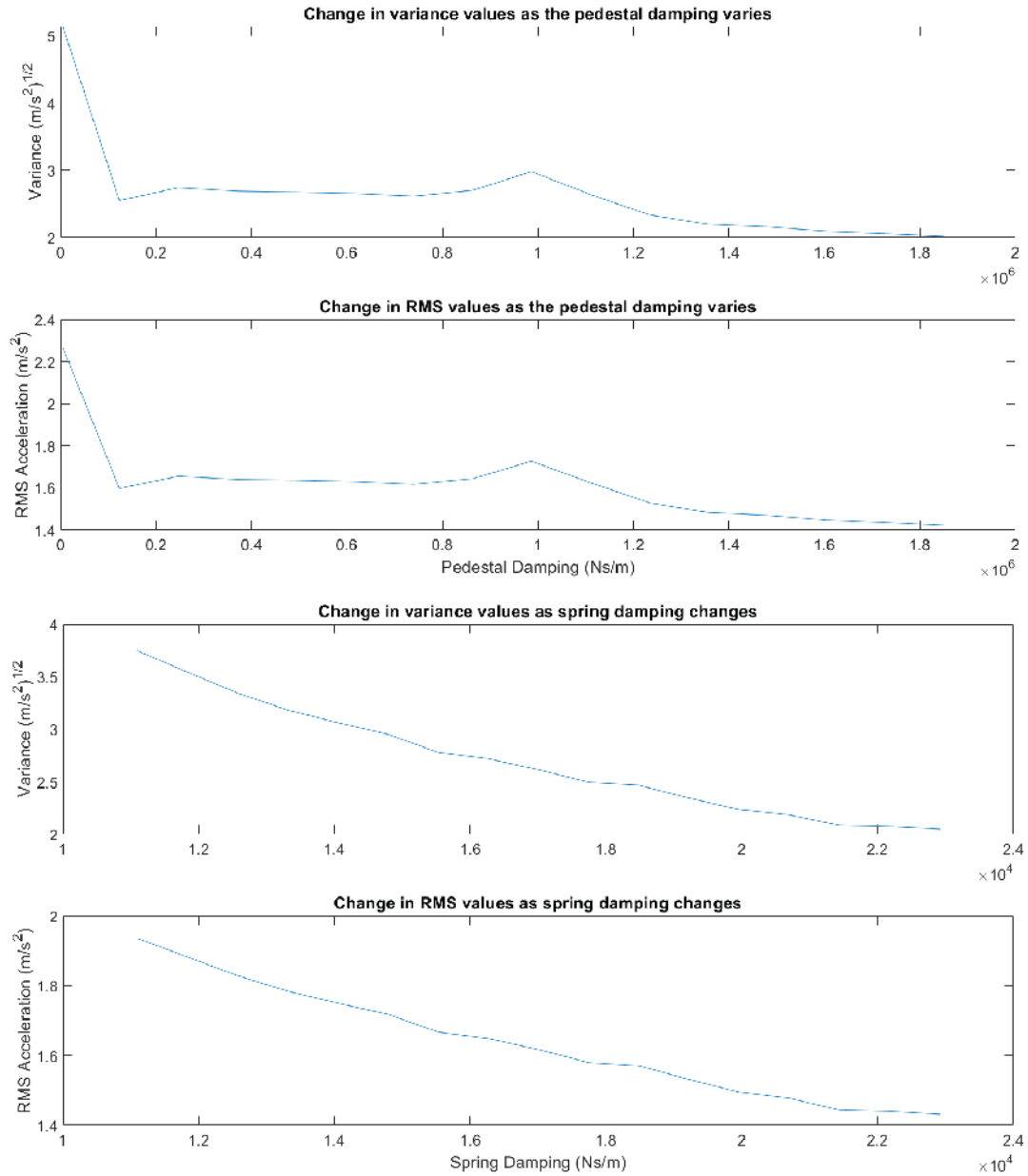
Figure 5.23: Changes in the RMS and variance as the pedestal damping varies (spring damping constant) and the spring damping varies (pedestal damping constant)

more to the expected results with the change in the time-series data. The sampling rate is closer to approximately 41.5 kS/s during the simulation (error of less than 0.0425%). In the 1 mm defect data, the frequency representation is similar to the low-frequency data; however, it is shifted to the approximate range of the expected defect frequency. There are even approximate frequencies in the expected range of the defect (124 Hz and 180 Hz), where the true defect frequency is 152 Hz.
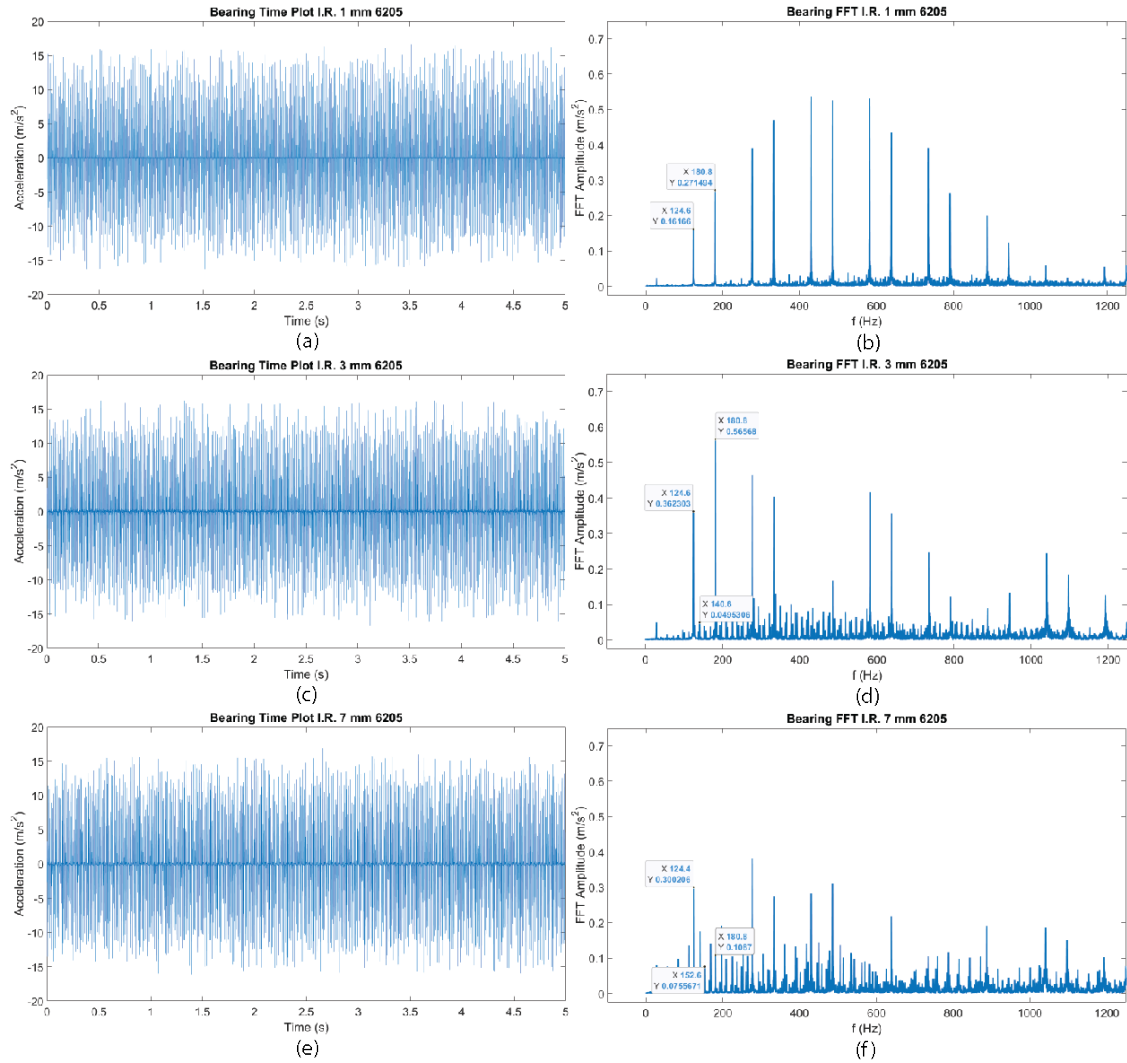
Figure 5.24: Time and FFT plots for the defect parameters when the model output is forced to the experimental data: (a) 6205 Time I.R. 1 mm defect, (b) 6205 FFT I.R. 1 mm defect, (c) 6205 Time I.R. 3 mm defect, (d) 6205 FFT I.R. 3 mm defect, (e) 6205 Time I.R. 7 mm defect, (f) 6205 FFT I.R. 7 mm defect

However, these are still off compared to the expected defect frequencies. The error in the sampling frequency may cause the difference in the frequency plot.

As the defect does progress, the bearing frequency representation does distribute and forms smaller peaks. With the 1 mm I.R. defect, the frequency peaks are distinct and separated. However, as the defect progresses, the frequency peaks begin to separate into more individualized peaks that are distributed along the frequency range. The peaks begin to increase around the defect frequencies. The defect frequency may be better captured with a larger defect. A peak in the 7 mm defect begins

258
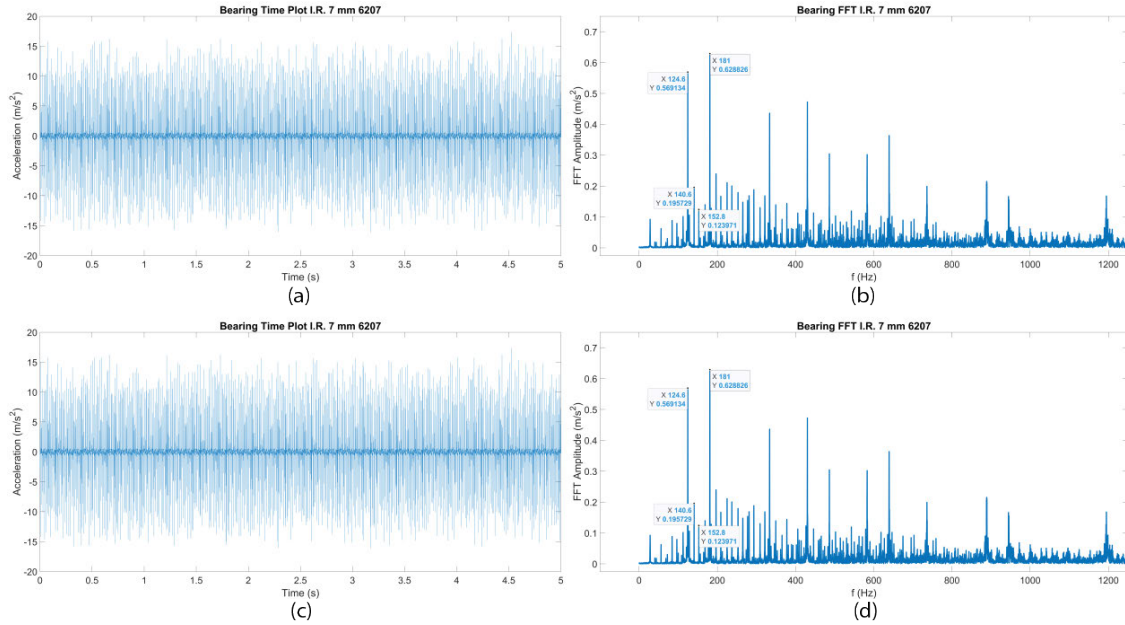
Figure 5.25: Comparison of the time and frequency plots between the 7 mm defects for the 6205 and 6207 bearings: (a) I.R. 7 mm defect 6205 Time Plot, (b) I.R. 7 mm defect 6205 FFT Plot, (c) I.R. 7 mm 6207 Time Plot, and (d) I.R. 6207 FFT Plot

to form around 152 Hz, although it is still much lower than the expected bearing amplitudes. By forcing the damping parameters, the bearing defect frequency begins to form closer to the expected values and supports the possibility of scaling data between different scenarios

Bearing data from the 6205, 6206, and 6207 1687 cases were considered to test the physics scaling process. The simulation results for the 6205 and 6207 cases are shown in Figure 5.25. An increase in the bearing amplitude is represented as an increase in the bearing data from the 6205 to the 6207 data. While the damping parameters were tuned to these different states, the increase in bearing data could also be represented based on the rolling element striking the defect due to the larger stiffness in the 6207 case. In this example, the bearing data from the 6205 represents the source case, whereas the bearing data from the 6207 case represents the target data case.

Theoretically, apart from the stiffness values and masses, the bearing components remain the same. The only components changing were the bearing geometry and the ball stiffness. Hence, in addition to the ground truth established in Figure 5.25 between the 6205 and 6207, a prediction was considered by using the damping parameters for either scenario to predict the vibrations of the corresponding bearing size. A test was considered where the 6205 bearings adopted the 6207 damping parameters and vice versa, and the bearing RMS was determined. In each instance, the

259

error when using the 6205 parameters to predict the 6207 vibration output was approximately 4%. The difference is measured based on the expected RMS values. The error when using the 6207 parameters to predict the 6205 vibration output was approximately 16%.

Since the damping parameters were kept constant in the simulation, the error could stem from the lack of adjustment seen in the stiffness values accordingly. The ball stiffness was scaled according to the difference in the rolling element relationships as determined in Section E.1. It is possible that a ratio could exist related to those different parameters that would allow for ease of prediction. As a quick matter of testing, a ratio was created based on the difference in the ball stiffness parameters and the inherent bearing deformation based on Hertzian contact stress. A power constant of 3/2 was incorporated based on the force relationship of the bearing. The reduction in error dropped the expected bearing response to approximately 1.25% for both cases. Table 5.14 contains the values for the different bearing predictions

Table 5.14: Predicted RMS values when using the opposite bearing parameters

Bearing Size	Expected	First Prediction	Second Prediction
6205	1.61 $\frac{m}{s^2}$	1.87 $\frac{m}{s^2}$	1.63 $\frac{m}{s^2}$
6207	1.97 $\frac{m}{s^2}$	1.68 $\frac{m}{s^2}$	1.96 $\frac{m}{s^2}$

Equation 5.34 shows the equation for modifying the damping parameters based on the application. In this instance, the example provides the 6207 bearing as the target value that is to be determined based on the known quantities from the 6205 and 6207 cases. The damping parameters are interchangeable in the application for this equation. In the example, the pedestal damping is referenced, but the stiffness equation could also work. δ_{HzC} represents the bearing contact stress for that particular bearing based on 5.17. Interestingly, the ratio derived is just the contact force without considering the defect. It does make sense that this parameter would be used as this is the inherent driving factor of force in the bearing within the bearing. It is possible that including an additional damping constant for the bearing force could further yield a better result.

$$R_{p,6207} = \frac{R_{p,6205} * K_{b,6205} * \delta_{HzC,6205}^{3/2}}{K_{b,6207} * \delta_{HzC,6207}^{3/2}} \tag{5.34}$$

Figure 5.26 shows the comparison of the actual 6205 model output and the predicted 6205 model output using the 6207 damping parameters and the same relationship for the 6207 case. The peaks between the two model responses for both bearing cases is similar and supports the error

found between Table 5.14. It is not certain whether the above Equation 5.34 transfers due to the similarity in the damping parameters versus the application of the test system. When testing the difference between the 6206 and 6207 case with Equation , the model response yielded an error of 9.3%. Hence, the scaling value does not hold true for every value.
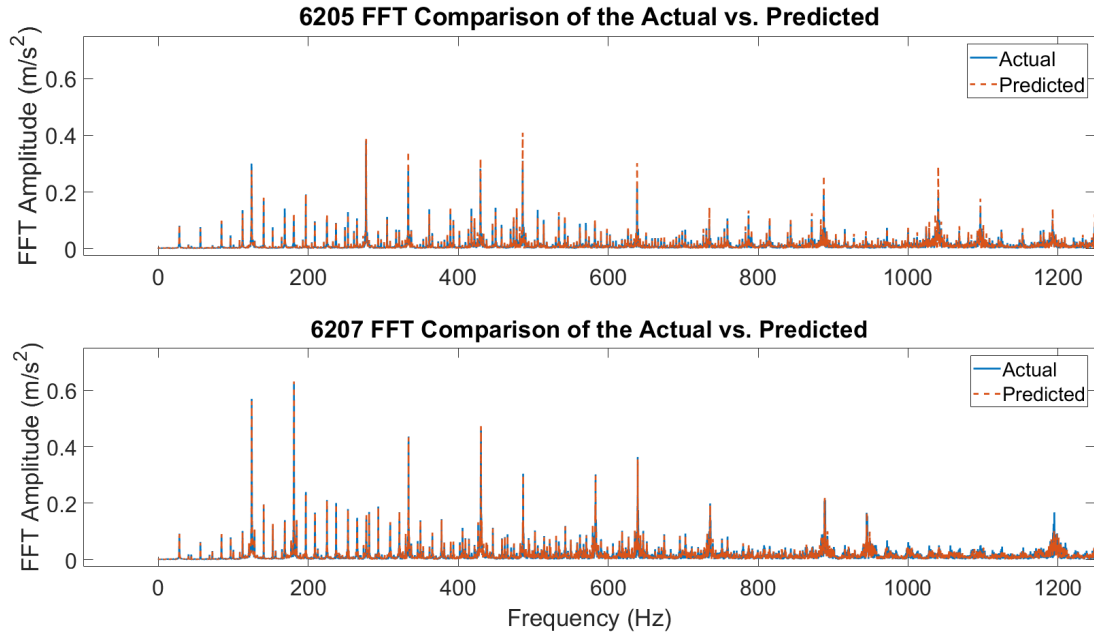
Figure 5.26: Actual vs. Predicted Model Output for the 6205 and 6207 bearing when using the scaled damping parameters

From the model predicted output, the physics scaling is tested in the following sections with the predicted value. The scaling relationship appears to hold true for the 6205 and 6207 cases. The 6206 case trend did not hold true in this relationship. However, it should be noted that the RMS output for the original 6206 case is similar to the 6207 case without modification. It is possible that the stiffness from the system plays a larger role in the determination. Table 5.15 shows the predicted values from the 6206 case where the non-scaled values (First Prediction) performed better than the scaled values used in the Second Prediction. Hence, further tuning is needed to dial in the scaling relationship between the different RMS values.

Table 5.15: Predicted RMS values when using the opposite bearing parameters for the 6206

Bearing Size	Expected	First Prediction	Second Prediction
6206	$1.96 \frac{m}{s^2}$	$1.91 \frac{m}{s^2}$	$1.77 \frac{m}{s^2}$

Despite the model converging in the time-domain based on the RMS and variance values,

it should be noted that some variation is expected in the frequency domain between the model actual values, the model predicted values, and the experimental data. Figure 5.27 demonstrates the difference in frequency response between the model actual, and model predicted values using scaled parameters from the 6207 case. Despite overlapping peaks and responses in the frequency domain from Figure 5.26, there is a noticeable difference between several peaks. However, when calculating the RMSE value, the value was still relatively low (0.0032). It is expected that this would change when compared to the experimental data. However, the convergence in the time domain does lead to the belief that there is some system similarity between the model and experimental data. It is possible that further improvement of the frequency response could consider manipulation of the stiffness and damping parameters as a lumped constant, as these would determine the time constant experienced by the system.
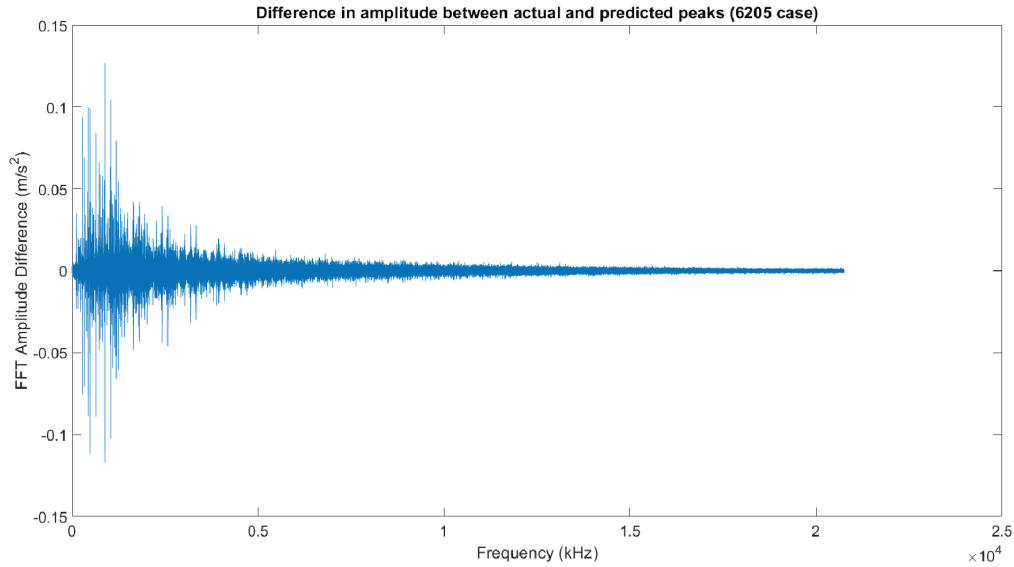
Figure 5.27: Error difference in the FFT peaks between the actual model values based on the FFT from the model tuned using the 6205 experimental data and the predicted FFT from the parameters scaled from the 6207 case

5.1.8 Physics Scaling - Test System

Due to the similarities between bearing 6205 and 6207 data, the data from these two sizes are used to test the overall process. The physics scaling process is the physics-driven process that utilizes the knowledge learned from the physics model to scale bearing data from one bearing size to

another bearing size. In either scenario, the time-series data can be used to scale the bearing data, or the frequency data could scale the bearing data. The frequency data is preferred to the time series data primarily due to the retention of physical information of the bearing due to knowledge of the defect frequencies and characteristic equipment. Hence, data manipulation occurs in the complex domain to ensure that it is possible to retain time series and frequency information.

Figure 5.28 shows the physics scaling process based on the different bearing groups. The source experimental data represents the data generated in an offline solution, whereas the target solution represents a different solution that needs the bearing data to fill a gap in the training history. From both solutions, there is respective knowledge of the bearing parameters from the physics model. The knowledge, represented by the dashed line, is passed to a labeled transfer function to take the bearing data from the source experimental data state to the target experimental data state. The generated IMF features are compared against each other for the augmented and true failure data, at least in the test stand application.
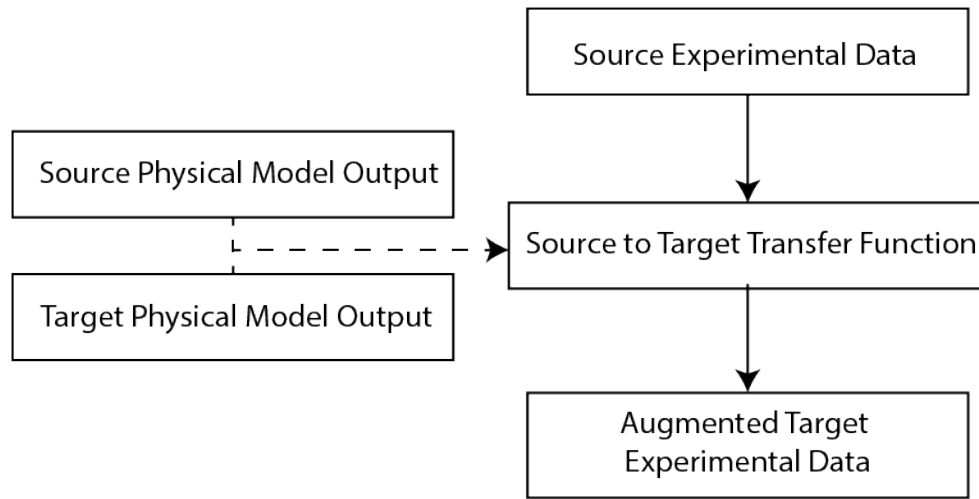
Figure 5.28: Bearing physics scaling process

There are two experiments to consider with the physics scaling process: Time-series manipulation and Frequency-space manipulation. Time-series manipulation involves limiting the data based on the output from the model and then scaling the time series values to the model output of the target using a ratio determined by the difference in RMS values. Frequency-space manipulation considers the creation of scaling values based on the complex domain of the FFT. The scaling values are applied along the length of the FFT and transform the source data to an augmented target state.

The Time-series manipulation process is outlined below:

i.) Three inputs enter the experimental transfer function: the source experimental data, the source model prediction, and the target model prediction.

ii.) The source data is scaled based on a ratio of the model predictions to change the source data to an augmented target data state.

iii.) The verification then occurs in the frequency space domain to determine if there are general similarities between the source and target domain.

During the physics scaling process using Frequency-Space Manipulation, the general procedure in the Source to Target Transfer Function is formulated as follows:

i.) Step 1: Three inputs enter the experimental transfer function: the source experimental data, the source model prediction, and the target model prediction.

ii.) Step 2: The source and target model predictions are taken by dividing the target by the source values at the individualized frequency locations. A reciprocal is taken to complete the operation because these values are represented in the complex domain.

iii.) Step 3: These scaling values are then applied to the experimental raw source data at each frequency point. However, due to the variation in the imaginary domain, the data were averaged between an upper and lower limit. The lower limit was defined based on applying a multiplication scaling factor, and the upper limit was defined using a divided scaling property.

iv.) Step 4: The bearing data are converted to IMF features based on the IMF bands from the data generation phase. These IMF features are compared against the ground truth of the experimental data to determine if there are noticeable similarities or differences in the data as a result of a similarity score.

An additional experiment was considered where the data were modified to utilize addition and subtraction versus multiplication and division. However, the change of operation removes some of the scaling aspects. The frequency bands are determined based on the IMF data generation from the earlier data generation phase. In this instance, the label "IMF" references the frequency range where the total energy content of the FFT is analyzed versus the values extracted as an application of empirical mode decomposition (EMD). The bands are configured as the following:

264

i). Band 1: 5000 - 10000 Hz - IMF 1

ii). Band 2: 2000 - 5000 Hz - IMF 2

iii). Band 3: 1000 - 2000 Hz - IMF 3

iv). Band 4: 300 - 1000 Hz - IMF 4

v). Band 5: 100 - 300 Hz - IMF 5

vi). Band 6: 40 - 100 Hz - IMF 6

vii). Band 7: 20 - 40 Hz - IMF 7

The energy content underneath the bearing FFT curve calculates the features. The ground truth and the augmented values are calculated for each case and then compared against each other visually. Two different statistical tests were considered: Kruskal Wallis test and Levene's test. The Kruskal-Wallis test is a non-parametric test, testing the null hypothesis that the data from two different groups comes from the same distribution. Levene's test is a non-parameter test that assesses the null hypothesis of whether the data from two different groups has the same variances. Kruskal Wallis test is chosen due to the comparison of sample distributions and the lack of normality assumption. Levene's test is used instead of Barlett's test due to having less sensitivity to normality.

Figure 5.29 contains the distribution plot comparison between data that are not scaled prior to the augmentation process and data that are scaled before the augmentation process. In both distributions, the bearing data follow the same progression of data. The source data has a larger peak and smaller base, whereas the target data has a larger base and a smaller peak. That is expected to some extent based on the amplitude of variables. The augmented data appears as a medium between the different data types; having a wider base and a slightly greater peak than the target data.

The primary difference between scaling the time series data between the RMS values was to try and remove undesirable outliers. The removed data were filtered to 0; hence, the much larger amplitude at 0. However, the deviation in the scaled data does not appear to change the data progressions. The shape changes due to the increased amount of zeros in the data.

Figure 5.30 contains the FFT response for the data that were scaled (a) and the data that were not scaled (b) based on the RMS value. In truth, there was not a noticeable change in the
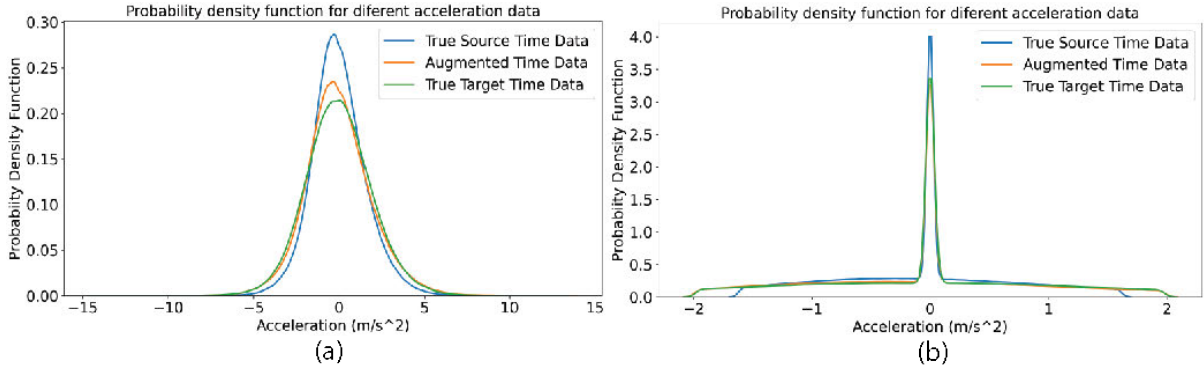
Figure 5.29: Time Scaled: Time series distribution 6205 Source, 6207 Target, and 6207 data predicted based on the model for data that was (a) not scaled and (b) scaled to the RMS value

frequency response as was expected based on the change in the time series data. For the difference in the original data and the augmented data, the overall values were increased at each frequency location; however, that as the only noticeable change for the bearing response.
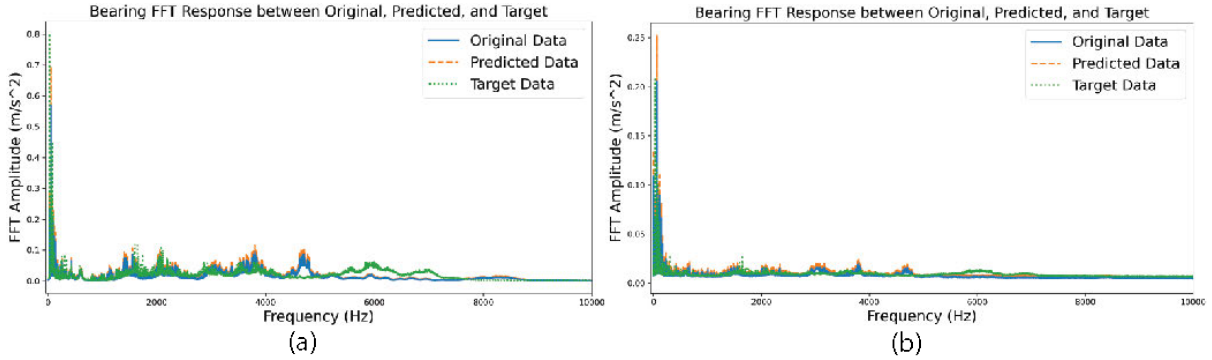
Figure 5.30: Time Scaled: FFT for 6205 Source, 6207 Target, and 6207 data predicted based on the model for data that was (a) not scaled and (b) scaled to the RMS value

Figure 5.31 shows the IMF feature comparison between the different data types for the data that were scaled (a) and the data that were not scaled based on the RMS value. Based on the observation of the values, IMF 3, IMF 4, and IMF 6, appear to show the most similar results in terms of the range for the data not scaled. For the data that were scaled, only the IMF 3 case were closer to each other. However, for the IMF 1 and IMF 7 case, there appeared to be a closer convergence in the data. However, that could have been a result of the filtering based on the RMS value, moreso than a change in the augmentation technique.

Initially, a Kruskall-Wallis test was considered to see if the means from the IMF features
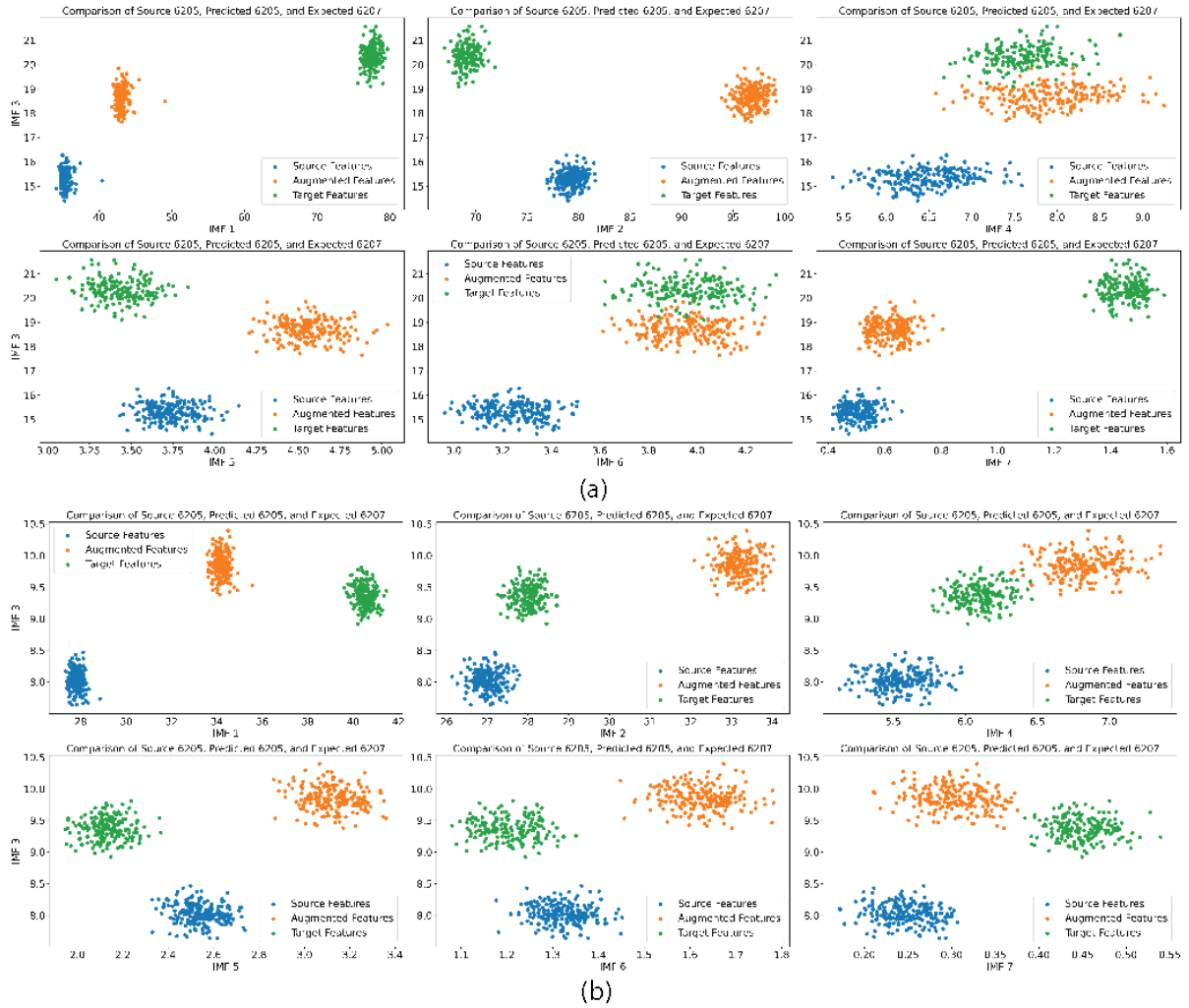
Figure 5.31: Time Scaled: IMF Features for 6205 Source, 6207 Target, and 6207 data predicted based on the model for data that was (a) not scaled and (b) scaled to the RMS value

from the different data types were statistically similar. However, there was not a significant result to state that the IMF features were similar. Levene's test was considered to see if the variances were statistically similar. There were two IMF features that were similar in the not scaled data between the augmented and target features, whereas the there were four IMF features that were statistically significant between the augmented and target features. Table 5.16 contains the p-values for the different statistical tests. It should be noted that there are similarities between the augmented and source data as well. However, with the augmentation, there is some level of similarity between the augmented and target features. It possibly that some of the bearing physics relationship is carried over as a result of the data transfer.

267

Table 5.16: Levene's test p-values for the different IMF features from time-series augmentation

Data Comparison	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7
Augmented - Target (Not Scaled)	0.008	0.006	0.4831	5.16e-5	4.82e-4	0.267	0.008
Augmented - Source (Not Scaled)	0.005	0.004	0.014	0.008	0.008	0.010	0.003
Augmented - Target (Scaled)	1.88e-4	6.97e-5	0.617	3.51e-6	0.811	0.809	0.407
Augmented - Source (Scaled)	0.067	0.082	0.011	0.015	0.019	0.005	0.0013
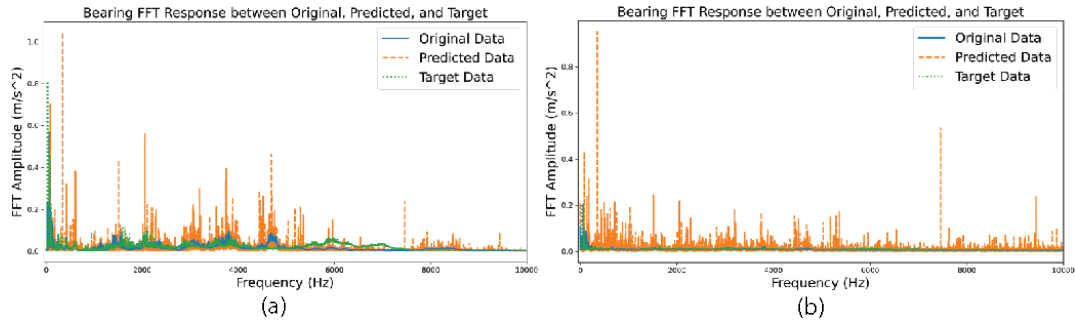
Figure 5.32: Frequency Scaled: FFT for 6205 Source, 6207 Target, and 6207 data predicted based on the model for data that was (a) not scaled and (b) scaled to the RMS value

Figure 5.32 shows the frequency scaling for the physics-scaling to show the difference in the IMF features for the (a) scaled and (b) not scaled to the RMS value. In this instance, the frequency plot appears to greatly outperform the bearing data in terms of the relevant amplitude. The primary deviation comes from the difference in the model output as seen in Figure 5.26. The scaling of the individual values greatly increases the over all expected output of the bearing response. It is also possible that the slight deviations in peaks contribute in the individual scaling value. Efforts were made to limit this output during the augmentation process such as adding in the RMS scaling; however, that did not appear to influence the result substantially.

Figure 5.33 shows the difference in the IMF feature plots when the data are scaled (a) and (b) not scaled. When the data is not scaled, IMF 2 and 4 appear to have similar properties. For when the data is scaled, none of the IMF feature groups appear to have similar properties. The bulk of the deviation may stem from the large difference in the frequency response compared to Figure
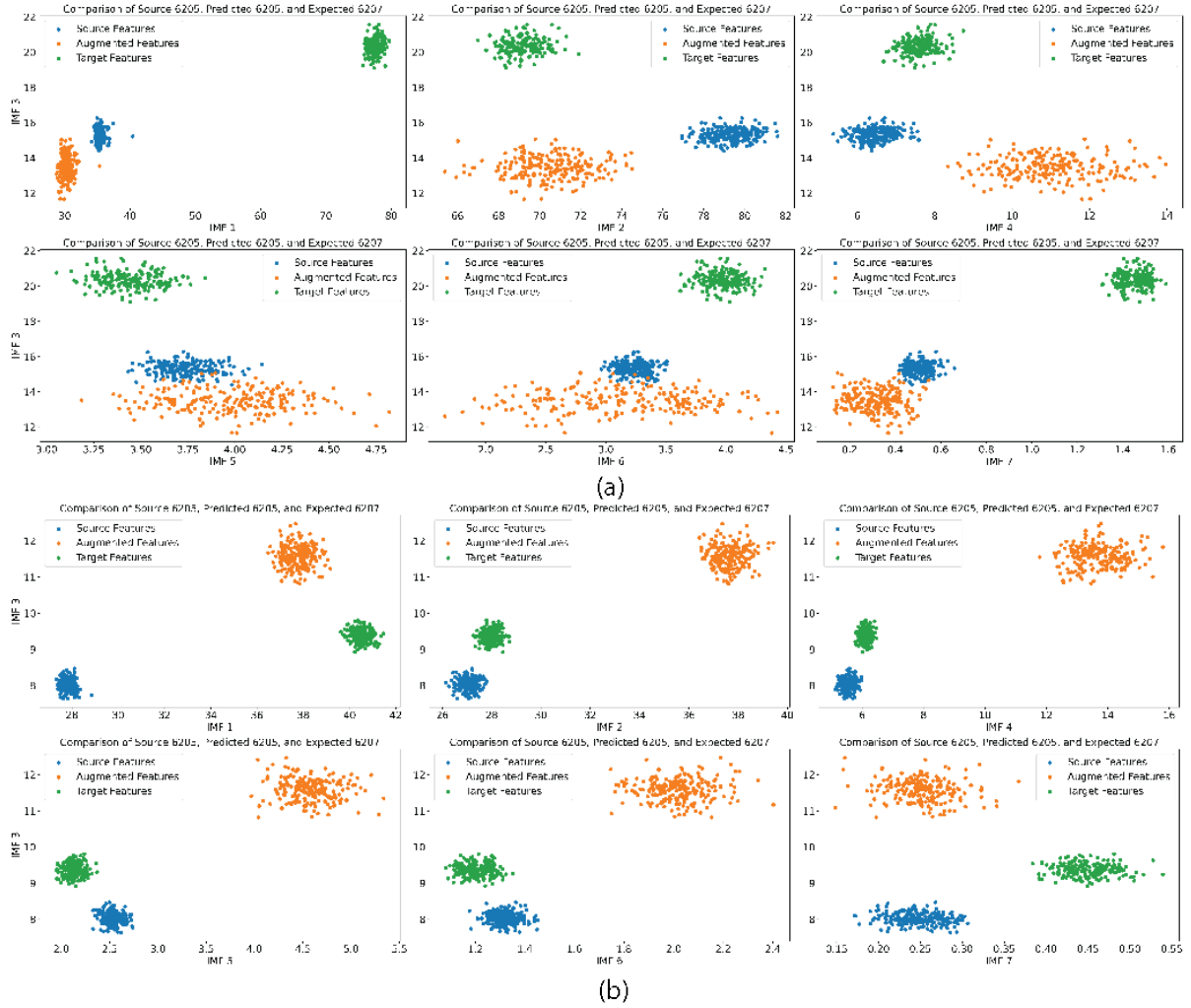
Figure 5.33: Frequency Scaled: IMF Features for 6205 Source, 6207 Target, and 6207 data predicted based on the model for data that was (a) not scaled and (b) scaled to the RMS value

5.32. More manipulation may be needed in the frequency domain to verify and adjust the physics scaling results to account for large shifts or deviations in frequency between the model responses.

Table 5.17 contains the different p-values for the statistical tests run over the different augmented data when scaled and not scaled by the RMS values. In this instance, the statistical values support that none of the values have an equal variance between the different feature sets. Directly manipulating the frequency space may cause an unexpected deviation from the anticipated results. Several limiting factors, such as a maximum amplitude based on the model output, were considered. However, that could be affected by components and vibration sources not anticipated by the model.

269

Table 5.17: Levene's test p-values for the different IMF features from frequency-space augmentation

Data Comparison	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7
Augmented - Target (Not Scaled)	0	0	0	0	0	0	0
Augmented - Source (Not Scaled)	0	0	0	0	0	0	1.45e-4
Augmented - Target (Scaled)	0	0	0	0	0	0	0.210
Augmented - Source (Scaled)	0	0	0	0	0	0	0.443

Based on the different scaling processes, the time-series augmentation process yielded data with similar characteristics in the different IMF groups. In the frequency space augmentation, the process appeared to struggle with the large deviations in the frequency response. The error could have stemmed from using multiplication and division as the main method for the augmentation. Addition and subtraction could be considered; however, that removes the scaling aspect of "stretching" an FFT or the residuals of an FFT to augment data to a different defect. The spectral augmentation is discussed in the following section (Section5.2.2) with more of the data-driven application of subtraction and addition.

5.1.9   Bearing Physics Discussion and Limitations

The method described outlines the process for setting up the pipeline to create a bearing physics model, determine parameters, and scale data between different bearing sizes and potentially dissimilar applications. The primary benefit of incorporating the physics model is verifying the bearing response from the experimental test stand and providing reasoning behind the possibility of scaling data between different applications. As was seen in the Chapter 4, the bearing data can vary greatly even when considering different bearings in near identical situations. Hence, the incorporation of a physics model can verify and constrain expected real world consideration. Utilizing that output model, predictions can be made that then support the understanding of what can occur with data scaling between other applications.

However, there are downsides to this method as well; primarily in how the scaling actually

occurs. As is evident based on the results from Section 5.1.4 and 5.1.5, even when constructing the model, there are noticeable deviations between the reality of configuring parameters and then tuning the expected parameters. In the idealized scenario where the parameters were derived based on theory (Section 5.1.4), the bearing data time-series response fell in line with expectations and provided basis for the expected results. However, the frequency domain did not conform to an expected response. Conversely, in the model response where the parameters were skewed to the experimental data (Section 5.1.5), the frequency results conformed closer to the expected model response. However, the time-series response does not meet the expectation with the real world response.

In response to the relation to a real-world physical system, one of the primary observations is the lack of knowledge surrounding such systems to outsiders. Of course, some information must remain proprietary and some universal assumptions can be made, such as material constants and kinematic information. However, these parameters are all inherently related through the dynamics of the system and can have significant impacts on modeling. In Chapter 5, additional exploration of the subject occurs by attempting to scale some of the test stand data to data from a potentially similar failure to a bearing in production motor, based on the universal assumptions.

To implement physics scaling on a broader application, further consideration should be provided on modeling other types of bearing defects such as contamination and lubrication fault propagation. Considerations could also be further given to change the defect shape and simulate shapes closer to the expected Hertz contact. Finally, another consideration could involve the use a decay function with the frequency amplitude to limit potentially significant shifts in the frequency output.

5.2 Physics-Informed Method: Spectral Augmentation

The proposed method for data-driven transfer is the spectral augmentation method. The technique is physics-informed because it uses physical analysis related to the bearing through the bearing fault frequencies to determine the necessary components for data transfer. The ensuing section discusses prior data augmentation methods for bearings and the difference between the data-driven and physics-informed measures. The differences are captured in the methodology and formulation before providing examples. The following sections cover the benefits and limitations of

these methods over the physics filtering/scaling process.

5.2.1   Vibration Data Augmentation: Summary

In the background section, an example of time splicing for vibration data was provided in Figure 2.17. Other data augmentation techniques as discussed by Li *et al.* [363] and Yu *et al.* [364] involve adding Gaussian noise, masking noise, amplitude shifting/data zooming, time stretching, signal translation/ data reversing, and segment splicing. Figure 2.17 is an example of local time splicing. The following list describes the other methods briefly:

i.) Gaussian Noise data augmentation adds Gaussian noise to the vibration and audio signals. Gaussian noise is commonly found in signals as a result of sensor noise or a grounding error in the circuit. Adding in the noise creates new data that may experience noise addition due to faulty equipment in the real world. figure 5.34 provides an example of Gaussian noise added to a sensor signal.
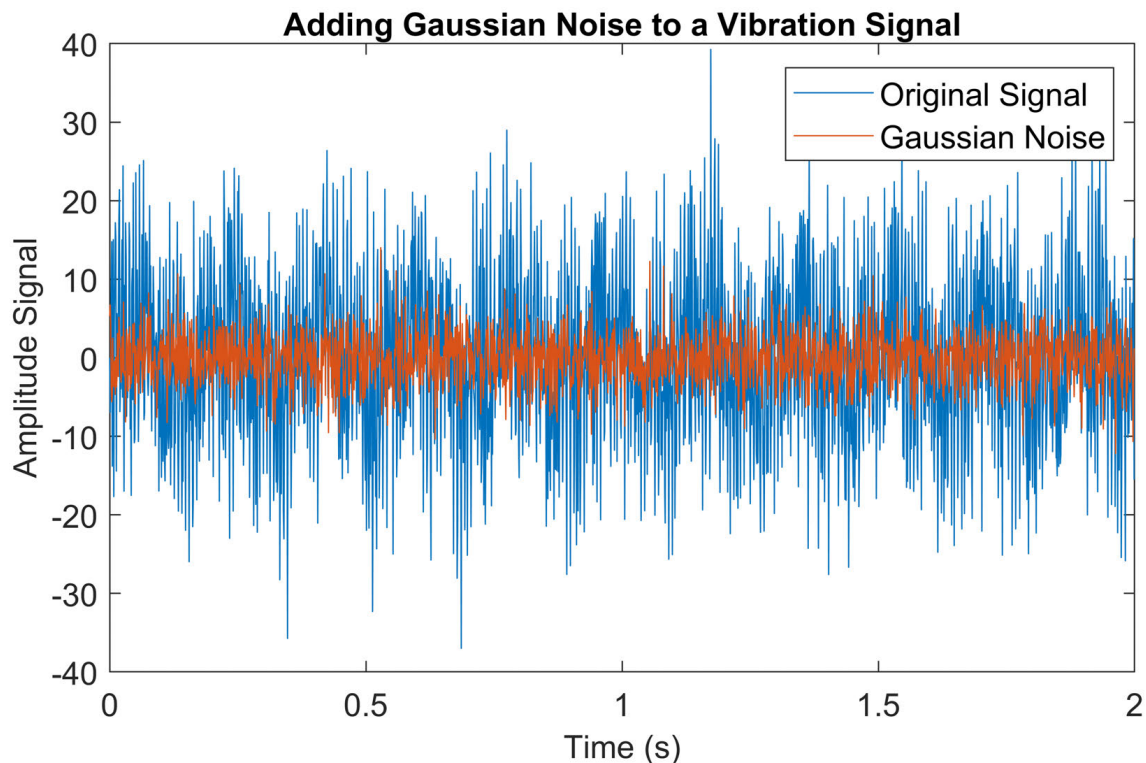
Figure 5.34: Example of Gaussian noise added to a generated vibration signal

ii.) Masking noise is a similar concept but applies random noise to the signal versus the Gaussian noise, which is a set noise generator.

iii.) Amplitude shifting [363] and data zooming [364] modify the vibration amplitude by scaling the data vertically by a known scaling factor. In the instances provided by both Li *et al.* and Yu *et al.*, the scaling factor is relative to the maximum data point within a segment. Yu *et al.* provides a selection of the data augmentation process relative to a certain amount of data described as local versus global scaling values. A local segment is a data segment from a large data class defined by a certain number of points. The global segment refers to the entire data class.

iv.) Time stretching considers stretching the vibration signal in the horizontal direction similarly to the amplitude shifting in the vertical direction. In this case, the entire data segment is stretched along the time axis until a predetermined point and is intermittently padded with zeros.

v.) Signal translation [363] and data reversing [364] seek to change the order of data by ordering the data in reverse order at either a local or global level. Figure 5.35 considers reversing a vibration signal based around a local segment.

vi.) The final data augmentation method considers the local segment splicing. In this method, the entire data segment is split into a number of different segments. These segments are then rearranged to create a new data segment. The reordered data represents a new dataset for that particular class.

Yu *et al.* [364] determined their data segment sizes based on the bearing characteristic frequencies. The segment size selection was to ensure that at least one impulse event (defect event) was captured in each bearing data class. Li *et al.* [363] divided the data based on the total availability and then applied multiple data augmentation techniques to the bearing data.

An area not covered, though, is the incorporation of frequency elements and physics parameters in these data augmentation techniques. Yu *et al.* considers the bearing frequency elements in selecting the data augmentation segments. However, in each of these applications, no new data is created as a result of the process. The segment splicing and data reversing represent processes that rearrange data, similar to oversampling. In this case, the data would reinforce certain trends

273
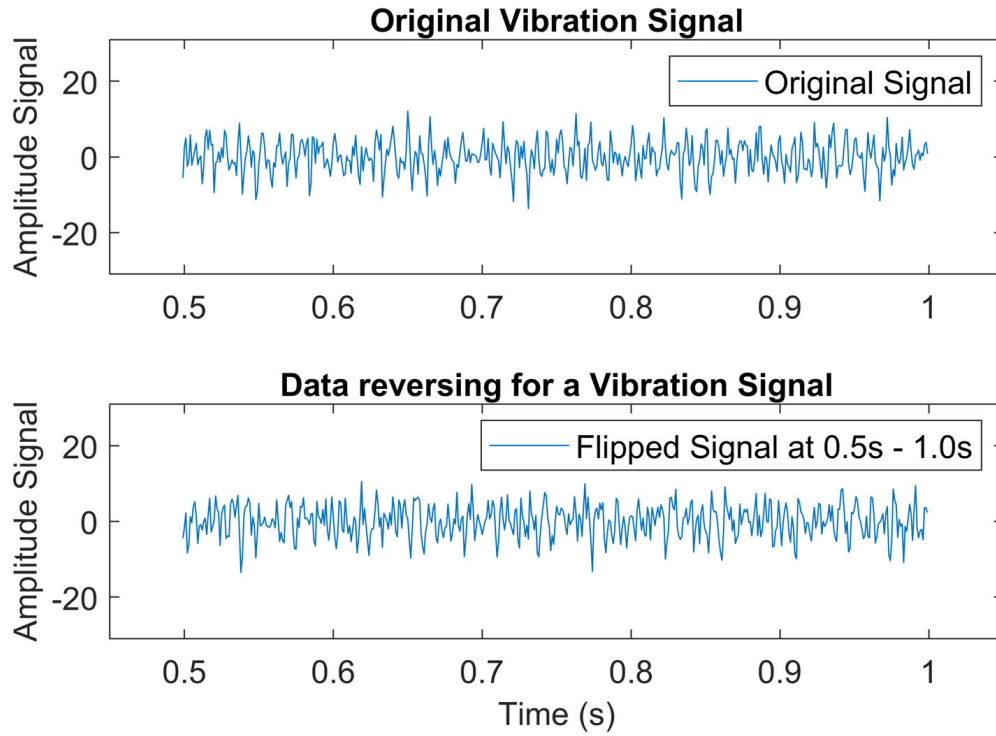
Figure 5.35: Example of signal reversing

through data augmentation rather than changing the sampling process. Figure 5.36 shows the difference in using data splicing for the raw data and the feature effect. Whereas the time features look similar, the effect on the frequency domain constitutes an entirely new signal not representative of the original signal.
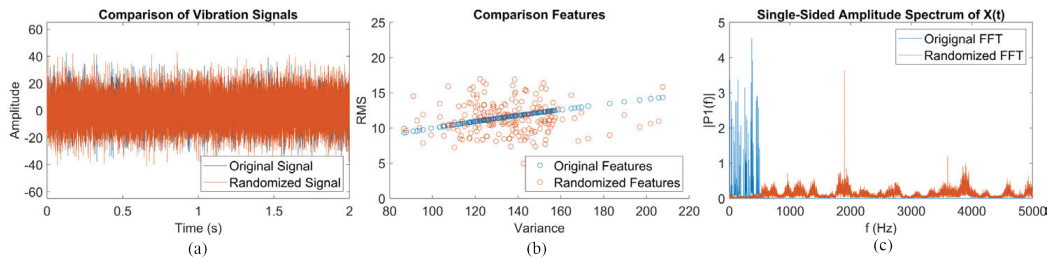
Figure 5.36: Example of the changing of the raw signal by segment splicing in (a) and the effect on the features in the time domain (b) and in the frequency domain (c)

As a result, a new physics-informed method is proposed to retain the expected frequency information for vibration condition monitoring while growing the range and usefulness of bearing

datasets. The physics-informed process is termed the Spectral Augmentation Process and combines the knowledge related to know frequency ranges for rotational equipment with the common methods employed with spectral augmentation. This method contrasts the Physics Scaling Process in that some failure data from the real system is needed to create the data. In that vein, the data serves more as a process to enhance an existing data set versus creating new data.

5.2.2 Spectral Augmentation Process

The spectral augmentation process combines spectral subtraction and data augmentation to create new data points, using a physics-informed process, rather than relying solely on a data-driven process. In the traditional application, spectral subtraction considers the removal of noise from a signal by subtracting the noise spectrum from a signal spectrum [524]. Noise signals generally comprise background noise that attempt to hide or obscure the expected diagnosis. Hence, the process of spectral subtraction usually involves identifying a signal (Y) that has both noise elements (N) and the desired signal (S) as shown in Equation 5.35. The equation is modified, as seen in in Equation 5.36, to reflect the spectral subtraction process. Figure 5.37 shows the effect of spectral subtraction for removing undesirable elements from the desired signal. In this instance, the noise elements are removed using their frequency transformation from the signal with noise leaving only the desired signal for analysis.

$$Y = N + S \tag{5.35}$$

$$S = Y - N \tag{5.36}$$

Spectral augmentation considers the spectral subtraction process as the initial step. However, instead of removing noise elements, the subtraction identifies the failure components from a bearing signal. For this formulation, the theoretical healthy vibration signal (V_{sh}) comprises both noise (N) and healthy (H) components as shown in Equation 5.37. The theoretical failure vibration signal (V_{sf}) comprises noise (N), healthy (H), and failure (F) components as shown in Equation 5.38. Following these formulations, only the failure components should remain from the bearing signal as shown in Equation 5.39. The extracted failure components are referred as "failure deltas", representing the isolated differences between the healthy and faulty vibration signal.
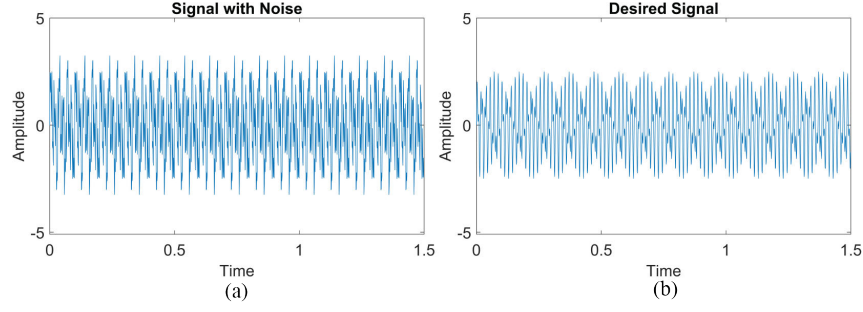
Figure 5.37: Spectral Subtraction to get a desired signal (S), shown in (b), from the baseline signal (Y), shown in (a).

$$V_{sh} = N + H \tag{5.37}$$

$$V_{sf} = N + H + F \tag{5.38}$$

$$F = V_{sf} - V_{sh} \tag{5.39}$$

Data augmentation now occurs by using the isolated failure delta by adding these isolated frequencies to create new failure data. The new failure data are created by taking existing healthy data and adding the failure deltas to the baseline signal as shown in Equation 5.40. By incorporating these failure deltas, the newly created failure data should contain similar aspects to the original failure data. The spectral augmentation process for vibration signals resembles image augmentation techniques where target images are superimposed into different backgrounds to increase the algorithm robustness. The process could be considered similar to prior augmentation methods involving noise injection into bearing data. However, the noise application can appear to be random at some times. As shown in Li *et al.*, the application of masking and Gaussian noise did not appear to positively increase the algorithm accuracy. By incorporating the spectral subtraction step, specific frequencies are targeted to generate new bearing data relative to the failure data.

$$AD_f = f_d + V_{sh} \tag{5.40}$$

Algorithm 1 provides the pseudo-code for the spectral augmentation process to create new

276

failure data. Different techniques are available for the frequency transformation based on the application [525]. For the spectral augmentation presented in this paper, a Fast Fourier Transform (FFT) [491] is used to convert the bearing time-series data to the frequency domain. Wavelet transforms were considered since they can perform with better accuracy in non-stationary applications [526]. This initial analysis does not consider non-stationary values; hence, FFT is considered sufficient for the time series to frequency spectrum transformation. As a means to extend this method to non-stationary applications, wavelet transforms, or Hilbert-Huang transform [488] could isolate the changing failure components in different frequency bands to perform the spectral augmentation process.

Algorithm 1: SpectralAugmentation(V_{sf}, V_{sh})

Inputs: V_{sf} - FFT failure data, V_{sh} - FFT baseline data
Output: AD_f - Augmented FFT failure data
split V_{sf} into training and test sets $\rightarrow train\_V_{sf}, test\_V_{sf}$
split V_{sh} into training and test sets $\rightarrow train\_V_{sh}, test\_V_{sh}$
Compute the average FFT for the failure data using $train\_V_{sf} \rightarrow avgFFT_{sf}$
Compute the average FFT for the baseline data using $train\_V_{sh} \rightarrow avgFFT_{sh}$
Extract the failure characteristics $avgFFT_{sf}$ - $avgFFT_{sh} \rightarrow f_d$
Compute the AD_f by using the $V_{sh} + f_d$

Due to the FFT limitations with speed-varying data in this formulation, only the steady-state data are considered for the initial spectral augmentation formulation. As a consideration to the speed-varying cases, the representative IMFs could be used to isolate failure changes as the signal changes. Despite the gradual changes called out in the test procedure, some segments of the IMFs remain steady, as shown in Figure 4.35. Hence, despite these speed changes, it could be possible to pick out the IMFs and provided targeted spectral subtraction to select frequency bands.

5.2.3 Spectral Augmentation Evaluation: Baseline Variation

Figure 5.38 demonstrates the extracted spectral signals between selected baseline bearings to show the possible variation in run-to-run operations. In a perfect scenario, there would not be any variation between the two different bearing signals. However, the variation experienced comes from unanticipated noise operating around the system. Part of that additional noise could come from the changing speeds over time. However, the larger spikes that appear at the operating speeds are unfiltered system noise from an outside source, possibly other operating equipment in the vicinity. The difference in low-frequency operating equipment could account for the shift in peaks within the different figures. The shift in peaks is captured by both the negative and positive peaks.
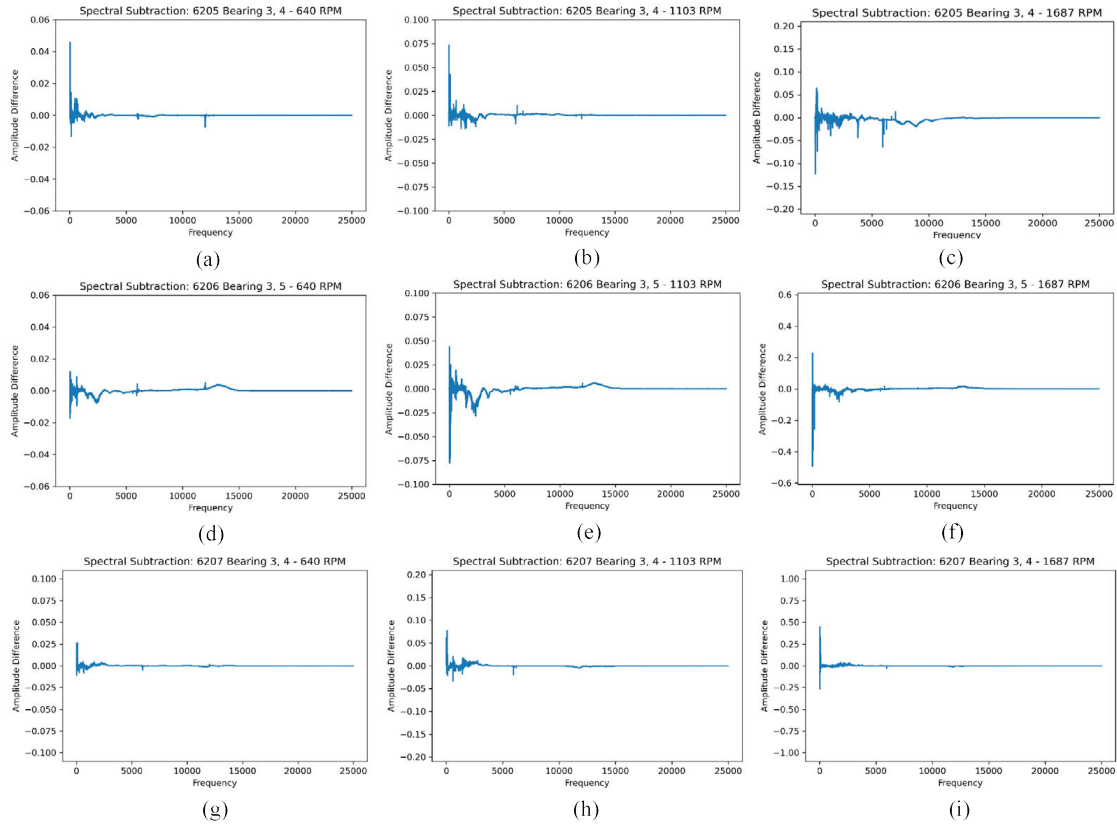
Figure 5.38: Spectral failure deltas for the comparison between different baseline bearings. (a) - (c) is the 6205 bearing at changing speeds, (d) - (f) is the 6206 bearing at changing speeds, and (g) - (i) is the 6207 bearing at changing speeds.

Figure 5.39 zooms in onto the peak difference between the different bearings for a collected group of bearings where the speed frequencies are not considered. For each bearing size, the variation appears dependent on what the minuend[3] and subtrahend[4]. For example, in the 6205 bearing data (Figure 5.39 (a) - (c)), as the minuend and subtrahend changes, the peaks appear to change as well. It should be noticed that the peak positions does not appear to vary as much in relation though. In the bearing 6206 case (Figure 5.39 (d) - (f)), the variation in peak height appears random in a case by case; the frequency position, though, is consistent. For the 6207 bearing (Figure 5.39 (g) - (i)), the bearing frequency position does not appear to match as cleanly in the 6205 and 6206 case. Another possibility for that difference in frequency position is due to slippage from the rotating elements. The motor slippage is calculated as 2.22% between the expected synchronous speed and

---

[3]Minuend is a real number from which another real number is to be subtracted
[4]Subtrahend is a real number subtracted from another real number

the rated speed for the system. Hence, due to that slippage factor, it is possible that the frequency shift seen in the 6207 case is due to the motor movement.
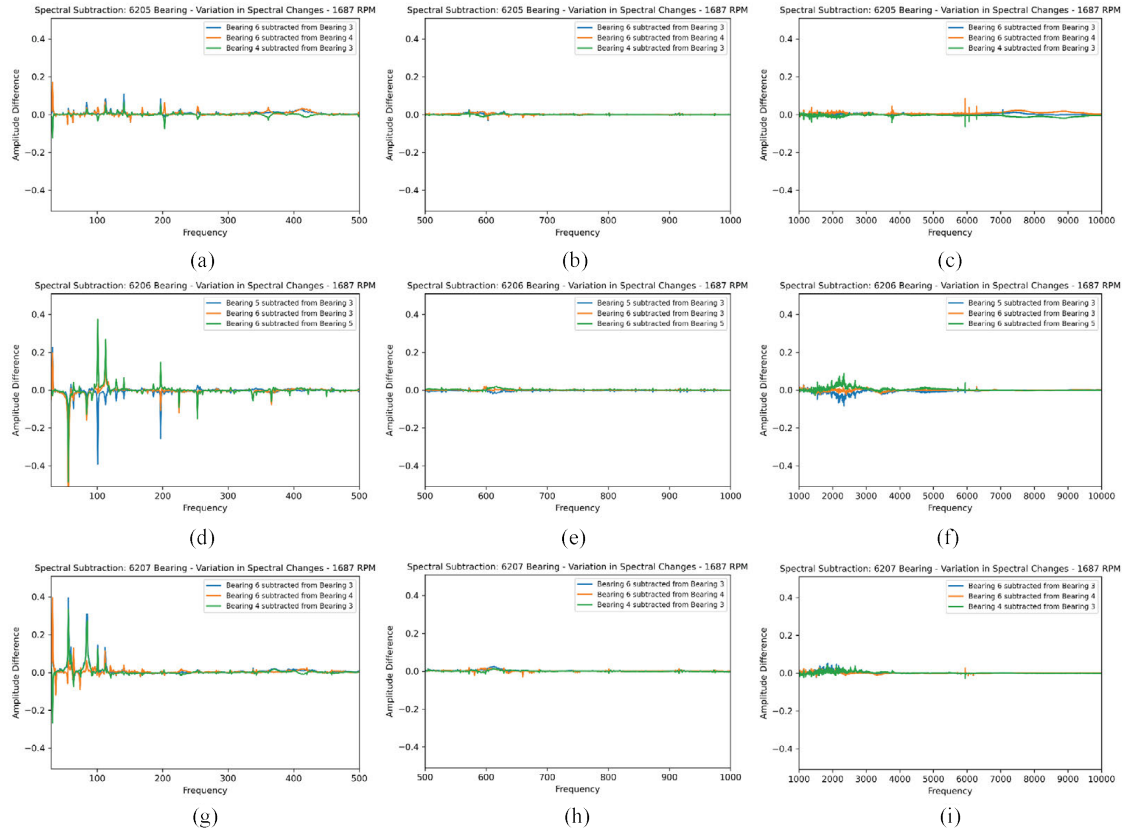
Figure 5.39: Variation in spectrum amplitudes for the 6205 bearing between (a) 30 - 500 Hz, (b) 500 - 1000 Hz, and (c) 1000 - 10000 Hz, the 6206 bearing between (d) 30 - 500 Hz, (e) 500 - 1000 Hz, and (f) 1000 - 10000 Hz, and the 6207 bearing between (g) 30 - 500 Hz, (h) 500 - 1000 Hz, and (i) 1000 - 10000 Hz.

The variation in the different spectral subtraction elements appears restricted primarily to the low-frequency bearing response. In assessing, the high frequency data, the variation between 500 Hz and 1000 Hz appears closer to 0. However, if one were to recall Figure 4.30 (details the average FFT response for each bearing size), not much variation was noted in this range. In the 1000 Hz - 10000 Hz range (Figure 4.30 (c), (f), (i)), the frequency positions appear to match in each bearing case. For the 6205 and 6206 case, though, the dependence on the minuend and subtrahend appears to affect the signs of the peaks.

Specific frequency ranges were considered to determine if there was any potential frequency overlap prior to the spectral subtraction method. The 6207 bearing was used as the example for

this method based around the frequency energy. The frequency energy is based around calculating the area under the FFT curve from the real portion of the FFT. Hence, the equation for calculating the frequency energy is based around Equation 5.41,

$$X_R(\omega) = \int_a^b x(\omega)d\omega \tag{5.41}$$

where $X_R(\omega)$ is the frequency energy for that portion of the FFT curve and $x(\omega)$ is the real component of the FFT. The upper (a) and lower bounds (b) are selected based around the upper and lower bounds for the different IMF bands as referenced during the Physics Scaling Process.

The basis for this calculation is to consider other FFT peaks outside of the expected frequencies. As noted in the in the frequency subtraction earlier, the amplitude may vary, but the position is not suppose to vary based on the geometric and kinematic expectations of the system. That is not always true, though, which is why an area calculation is used to capture and explain the different spectral subtraction. Figure 5.40 shows the different comparisons of the frequency energy with respect to IMF band 3. IMF band 3 was selected as the primary comparison due to the close similarity in the spectral subtraction from Figure 5.39 (i) for Bearings 3, 4, and 6. IMF bands 4 - 7 were not considered, mainly due to the large amount of variation seen in the low frequencies.
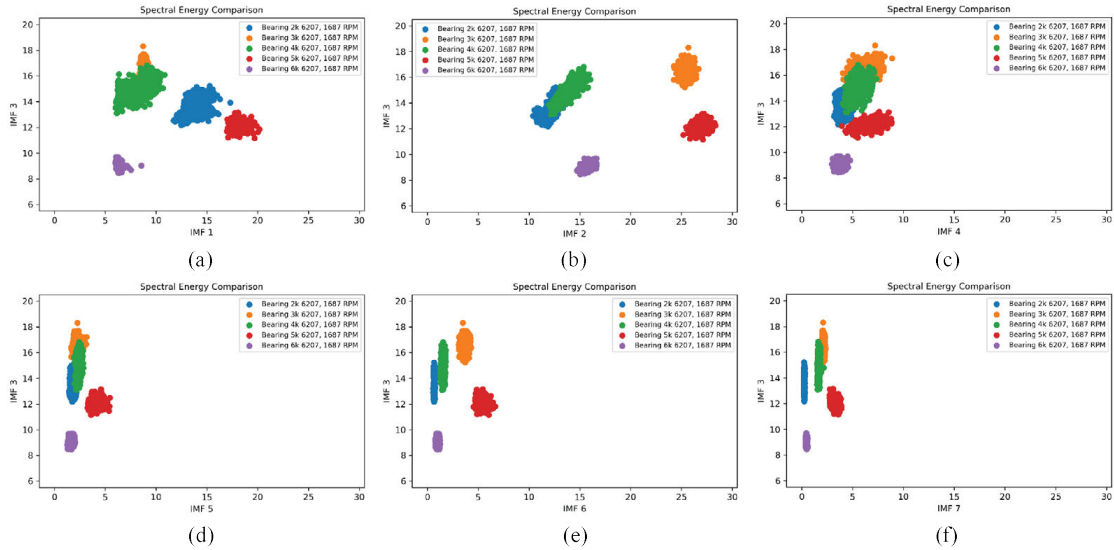
Figure 5.40: 6207 frequency energy based on IMF frequency bands for the 6207 bearing. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

Interestingly, though, the frequency energy for the IMF 7, 5, 4, and 6 had the smallest ranges across the different frequency representations: 3.71, 4.18, 5.9, and 6.11 respectively. It should be noted that the bands with more frequencies to consider had a higher range of frequency energy. However, despite that range, there are still separable clusters identified for each IMF band. The largest overlap based on Figure 5.40 occurs in (c) and (d), where the clusters are overlapped between Bearings 2, 3, 4, and 5. Bearing 6 is a deviation from the majority of the bearing clusters and could come from a reduced amplitude.

Hence, to create a homogeneous spectral failure delta for the different failure mode comparisons, the spectral failure delta for the 6207 bearing baseline was created by using an average spectral failure delta from the accumulated bearing data. A split was conducted of 80% training data to create an average spectral failure delta. The spectral deltas are created then using the average representation versus multiple individual representations for the spectral failure delta. The remaining 20% of the data, dubbed the test data, is then augmented using the spectral delta for a particular bearing with the average delta. Figure 5.41 provides the individual and average frequency response in the 6207 bearing (Figure 5.41 a) and the subsequent different variations of the spectral deltas (Figure 5.41 b). Th primary difference occurs around the initial spike at the speed frequency. The variation from the average response is approximately 95% at its maximum for the Bearing 5K. By conducting the transformation for the average to the failure data, the run-to-run characteristics for each bearing are retained without normalizing the total response. Currently, this response and method is valid for only bearings collected under the same geometric, kinematic, physical, and dynamic parameters.
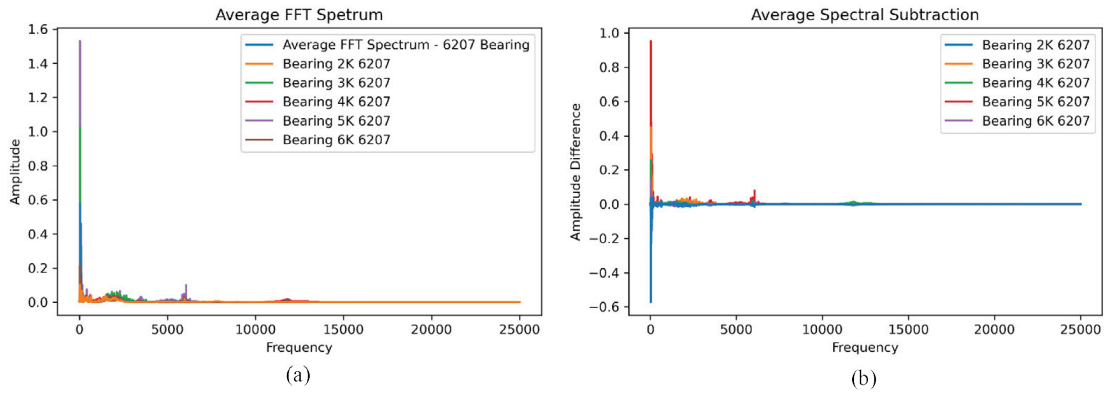
Figure 5.41: Average FFT representation for the 6207 bearing (a) and spectral subtraction (b)

Figure 5.42 compares the original IMF clusters to the test clusters for each bearing. The augmented test clusters form in the center of the original groupings. The data appear intermixed; however, the original data shape is retained in each case. For example, observe the Bearing 4K data depicted in the green data. The test data shape matches the original data shape, retaining the response. To determine if there was statistical similarity between the test augmented bearing data, a Kruskal-Wallis test was conducted to test if the data came from similar populations. Additionally, Levene's test was conducted to see if there was a similarity between the variances of the different independent sample groups.
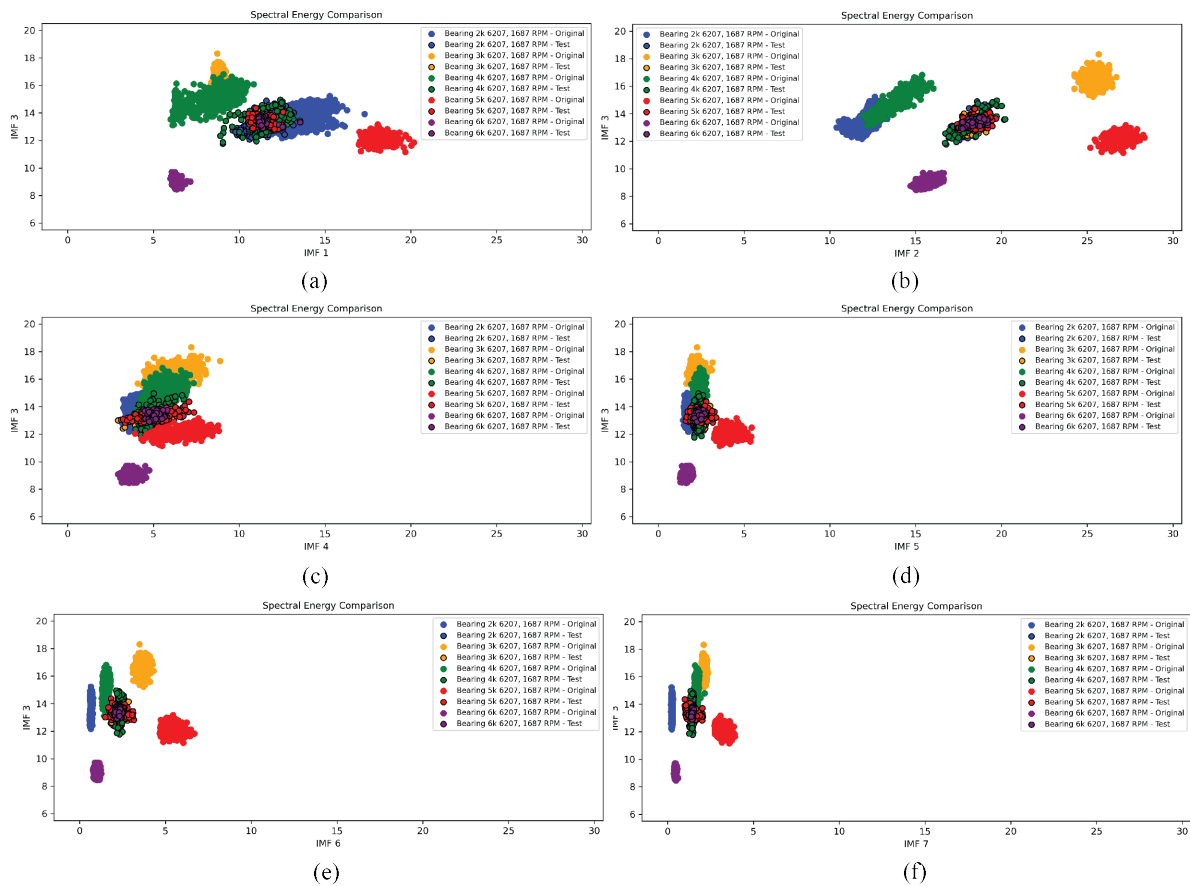
Figure 5.42: Train and test comparison based on the different IMF combinations for the 6207 bearings. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

Table 5.18 contains the p-values for the different statistical tests across the different IMF bands for the frequency energy. For the Kruskal-Wallis test, the IMF Band 1 rejects the hypothesis of similar means. On further analysis of individual bearing frequency energy means, Bearing 5 and 6

282

failed to reject the null hypothesis of the Kruskal-Wallis test. A possible reason could stem from the bearing noise centered around the speed frequency. The other IMF bands failed to reject the null hypothesis, implying that there were similar means between the different IMF groups. In the case of Levene's test, the variances were not similar in any of the group cases. The different variances could be due to the dissimilar original data shapes.

Table 5.18: Kruskal Wallis and Levene's test p-values for each IMF Range

IMF Ranges	IMF Band 1	IMF Band 2	IMF Band 3	IMF Band 4	IMF Band 5	IMF Band 6	IMF Band 7
Kruskal-Wallis Test	4.2E-7	0.13	0.14	0.48	0.75	0.80	0.40
Levene's Test	0	0	0	0	0	0	0

The conventional method considers only the real number variations from the signal converted into the frequency domain. The conversion of the time series data to the Fourier frequency space requires an absolute conversion to remove the imaginary terms. Theoretically, it is possible to perform the spectral subtraction calculation utilizing the raw spectral information [527]. Further exploration of the process concerning the raw FFT would incorporate a similar process as conducted by Zhang *et al.* [527]. However, since the focus for bearing condition monitoring is on the use of frequency information, this particular portion may be considered future work and is discussed further in the future work section of this dissertation.

5.2.4 Spectral Augmentation Evaluation: Failure Modes

5.2.4.1 Bearing Contamination Case: 6207

The initial formulation of the spectral augmentation was considered for the augmentation case of transferring data between bearings of similar sizes. In this instance, the IMF ranges are referencing the range of frequencies observed from the HHT and not the actual IMF values extracted by EMD. The data for each failure stage were split into "test" and "train" labels for the failure mode consideration for the spectral augmentation method. The analysis occurred in the frequency domain, normalized to the Fourier representation. The generalized baseline amplitude is used to represent each bearing class to extract the corresponding failure characteristics from each signal. In the initial formulation, the 6207 bearing demonstrates the differences and similarities between the healthy and

progressing failure classes.

Figure 5.43 shows the average FFT and the subsequent spectral difference between the average 6207 baseline case and the different contamination stages. From the spectral difference, there are key differences based on how damage changes from the initial contamination stage (Code 20) to the failure condition (Code 25). The Code 0 case is when uncontaminated grease is run through the bearing to see the particle's lasting effect on the bearing. Generally, as particles accumulated in the bearing, the higher frequency signatures increased, and some of the lower frequency ranges were greater than 100 Hz. For the Code 20 and Code 0 cases, the bearing frequency around the speed frequency was much lower than in some baseline configurations, possibly indicative of some noise or misalignment in the system. Interestingly, there appeared to be a critical point where the high-frequency response stabilized, as shown in Figure 5.44. Code 24 appears to be the peak in the high-frequency content (2500 - 10000 Hz).
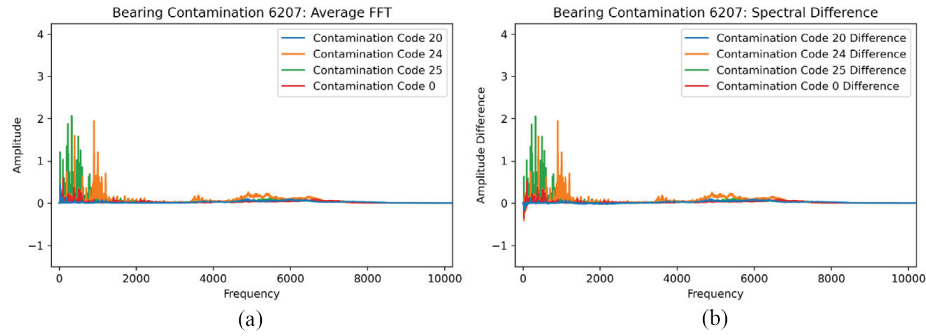
Figure 5.43: Contamination average FFT and the failure deltas. (a) 6207 average FFT, (b) 6207 spectral difference.

Figure 5.45 shows the different IMF cases for the bearing data augmented to the Code 20 case as compared to the real data case. There is a substantial increase for each IMF case depending on the different frequency bands assessed as compared to the baseline cases. For example, the range of the IMF 1 case has changed from between 5 and 20 in Figure 5.42 to the range of 145 to 180 as shown in Figure 5.45. The general increase is apparent over the corresponding IMF ranges, except in the IMF 7 range, where the comparison remains similar to the baseline scenario shown in Figure 5.45. The visualization of the IMF energy values characterizes the bearing baseline variance from the bearings of the same size. The overall distribution appears to remain the same despite the averaging of the results. It should be noted that for the lower frequency cases shown in IMF 6 and 7, some of the data appeared to become negative. Since, a negative normalized FFT is unlikely, an additional
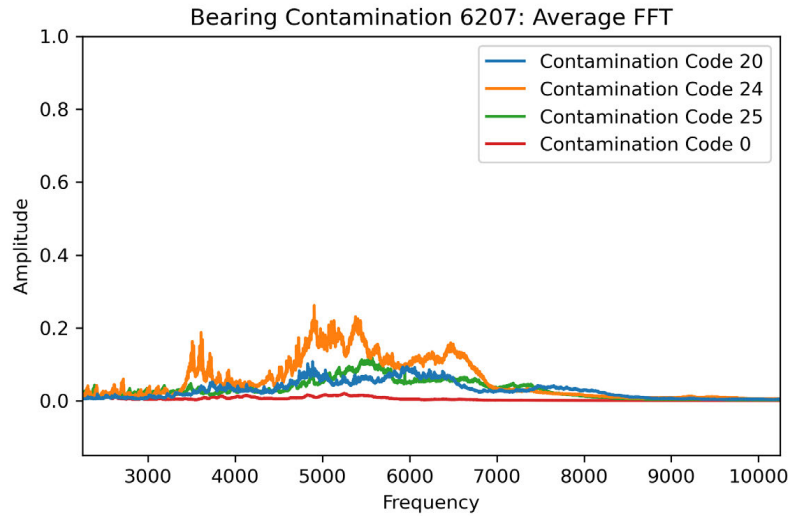
Figure 5.44: 6207 average FFT contamination response zoomed in on the higher frequency.

consideration may be needed to ensure that negative values are not considered in the algorithmic analysis.

Figure 5.46 shows the IMF cases for the bearing data augmented to the Code 24 case relative to the original data from the Code 24 contamination case. In this instance, the IMF ranges did increase across the different cases. The range appeared to decrease for the IMF 1 and 2 cases. However, the IMF 4 range increased by about 10x, signifying the shift in the frequency ranges most affected by the particle increase. The augmented contamination data appeared to fit this data similarly to the COde 20 case for the frequencies from the IMF 1 through 5 cases. It is possible that this stems from the fact that the contamination data experiences a noticeably higher variation due to the introduction of particles versus a known defect, such as in the fatigue case.The low-frequency cases of IMF 6 and 7 do not appear to change relative to the Code 20 case. This could be due to the lack of effect the particles appear to have on the low frequency vibration relative to the multiple strikes that occur each revolution.

The relative increase in contamination causes an increase in energy levels relative to the Code 20 data. However, the greater jumps in the IMF 4 and 5 cases could also be due to increased indentations relative to the Code 20 case. Another difference between Code 20 and Code 24 is the distinct data shape for each collection. In each case, there are three differentiable collections from the test. The highest amplitude collection represents the first collection case, the middle amplitude
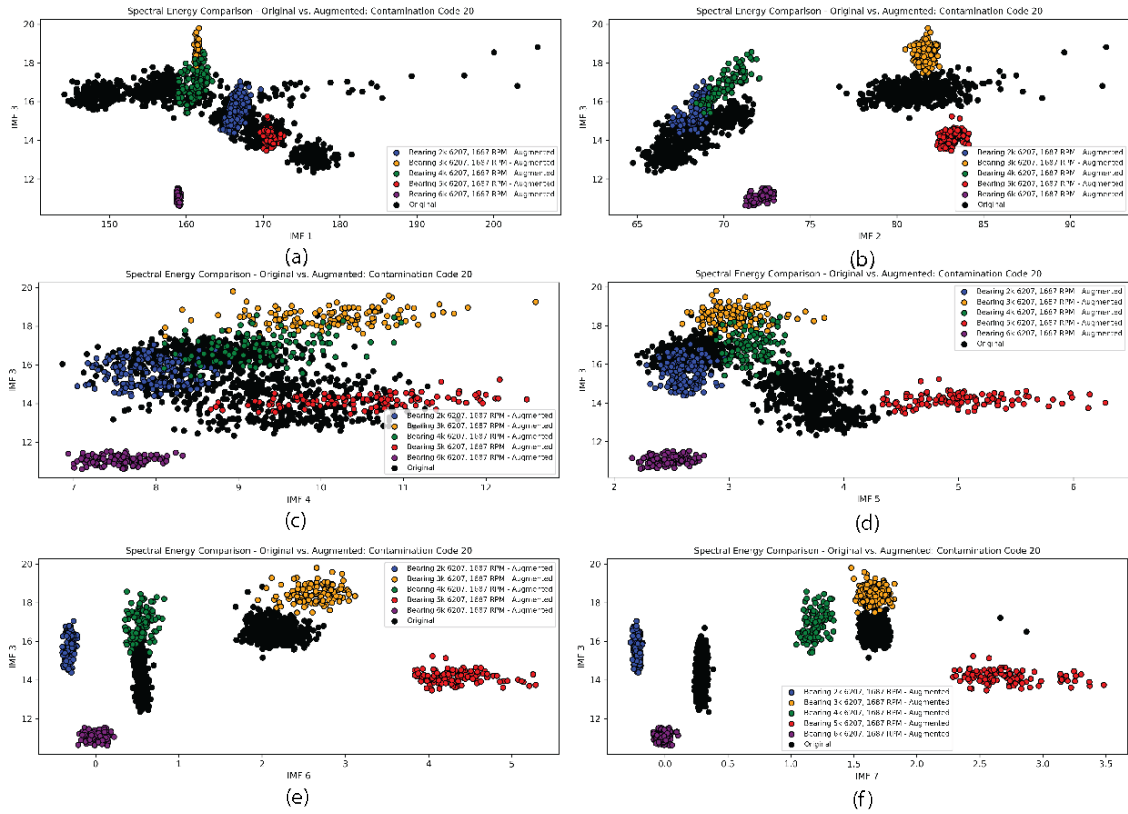
Figure 5.45: IMFs for the augmented bearing data for each 6207 case bearing versus the original bearing data in the contamination code 20. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

case is the second collection, and the lowest amplitude case is the third collection. One possible reason for the difference is the grinding of the contamination particles during the bearing operation. The particle breakdown may reduce the size of indentations caused on the bearing surface, thereby decreasing the vibration and subsequent energy levels.

Figure 5.47 shows the IMF cases for the bearing data augmented to the Code 25 case relative to the original data from the Code 25 contamination case. In this instance, the IMF 1, 2, and 3 cases decrease relative to the Code 24 cases. The data returns to a similar level as Code 20 for IMF 1 and 2. The IMF 3 range appears to decrease relative to the Code 24 case but remains greater than the Code 20 case. The IMF 4 and 5 values are greater than the Code 24 cases, signifying that the increased vibration appears to form closer to the defect frequencies. It is possible that the frequency of particle impacts reached a plateau that begins to form the expected indentations from continued contamination. This observation is further supported by the increased vibration energy for IMF 6

286
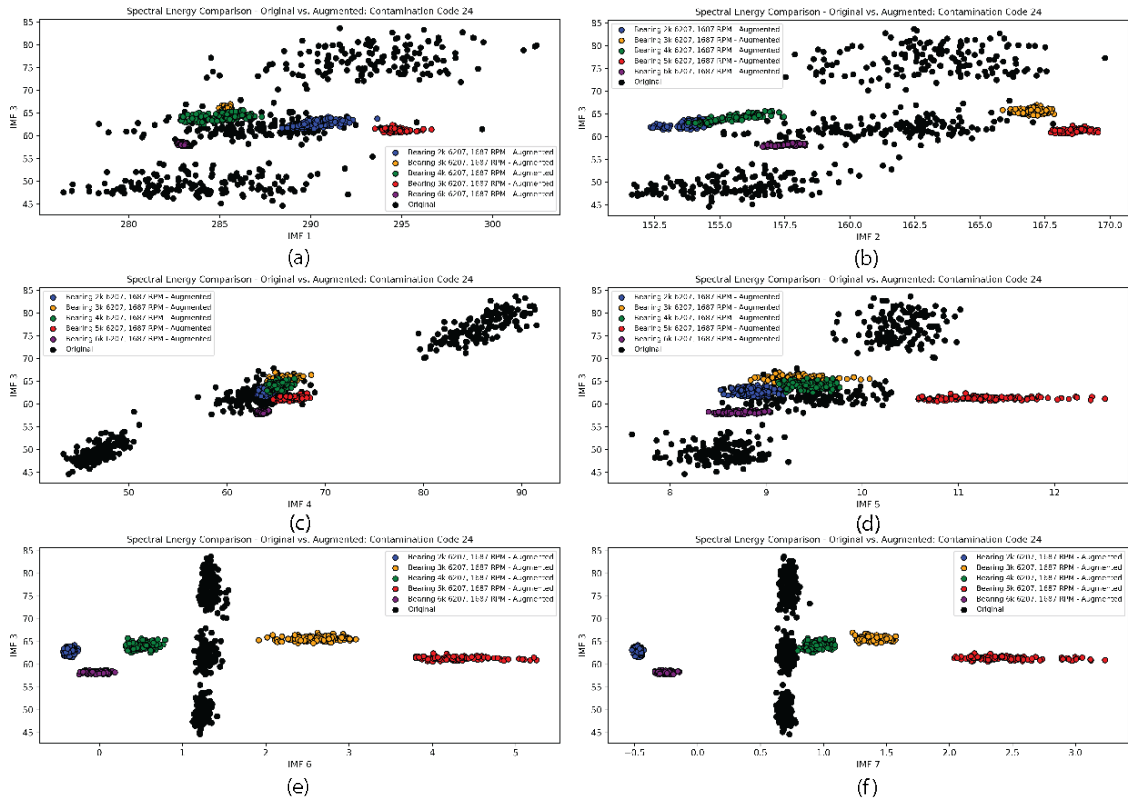
Figure 5.46: IMFs for the augmented bearing data for each 6207 case bearing versus the original bearing data in the contamination code 24. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

and 7 that corresponds more to the lower speed impacts. From the pictures provided in Section 4.3.2.2, the bearing surface represents a continuous surface of indentations triggered by the rolling of the balls across the bearing surface. The separation in the different energy levels could stem from a similar particle breakdown during the bearing operation.

Figure 5.48 shows the energy levels of the augmented data from the clean grease case after contamination compared to the real data from the same case. The first observation of the energy levels is that there is a single represented cluster instead of the multi-cluster formation as seen with the contamination cases with particles added. The variance in the baseline data is still noted as it forms around the different cases. Compared to the baseline characterization from Figure 5.42, there is permanent damage due to the introduction of bearing contamination. IMF 1 and 2 are lower than any of the contamination stages. IMF 3 remains greater than the Code 20 case but less than the Code 24 and 25 cases. IMF 4 is greater than the Code 20 case but less than the Code

287
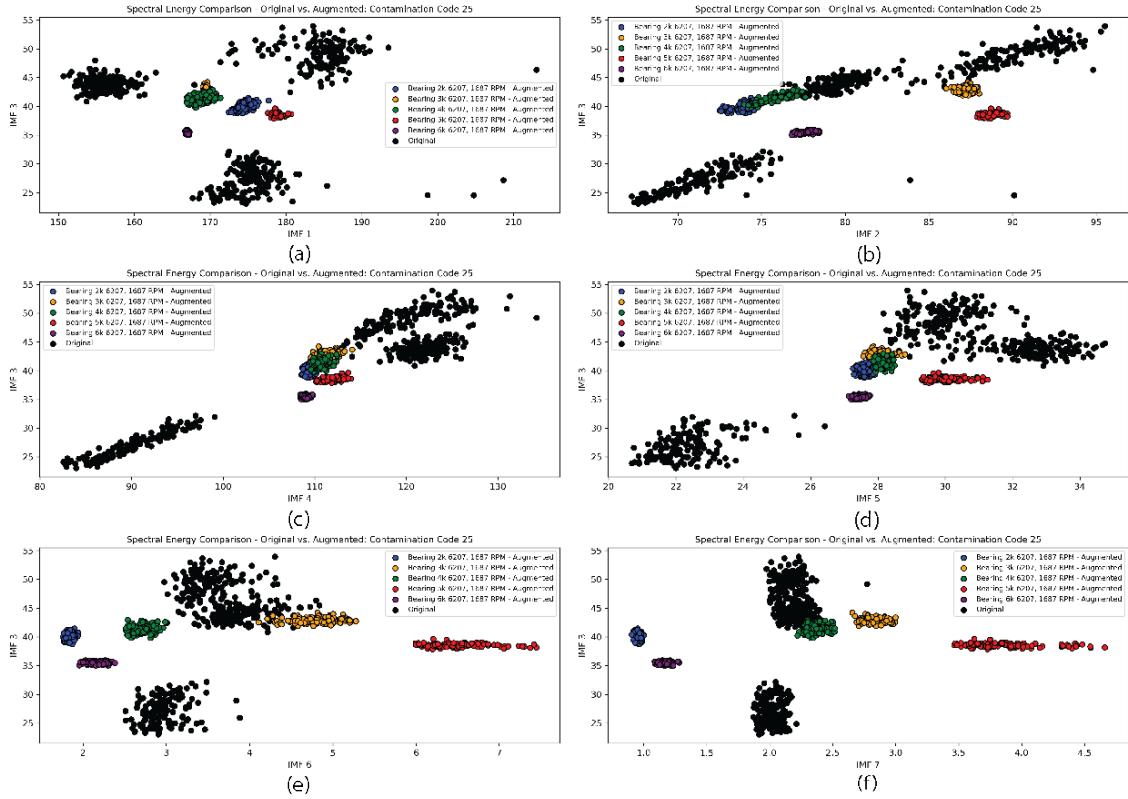
Figure 5.47: IMFs for the augmented bearing data for each 6207 case bearing versus the original bearing data in the contamination code 25. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

24 case. IMF 5 is similar to the Code 25 case. IMF 6 is similar to the Code 25 case, but IMF 7 resembles the Code 20 and 24 cases. Essentially, the Code 0 case represents a conglomeration of the different contamination frequency ranges. The bearing augmented features, as a result, appear to form around the real data. Hence, for the augmented case, the bearing data does appear to retain the same expected bearing baseline response despite having only a single cluster to consider.

5.2.4.2   Fatigue: 6207

Figure 5.49 shows the spectral failure deltas at the same scale as the contamination failure deltas. A possible reason for the lower response in the fatigue data as compared to the contamination data could stem from the additional impacts of a result from the contamination elements. The lower number of impacts causes a lower vibration response as a result. In this instance, the fatigue failure deltas are much lower and do not share the same expected consistency in growing as damage is

288
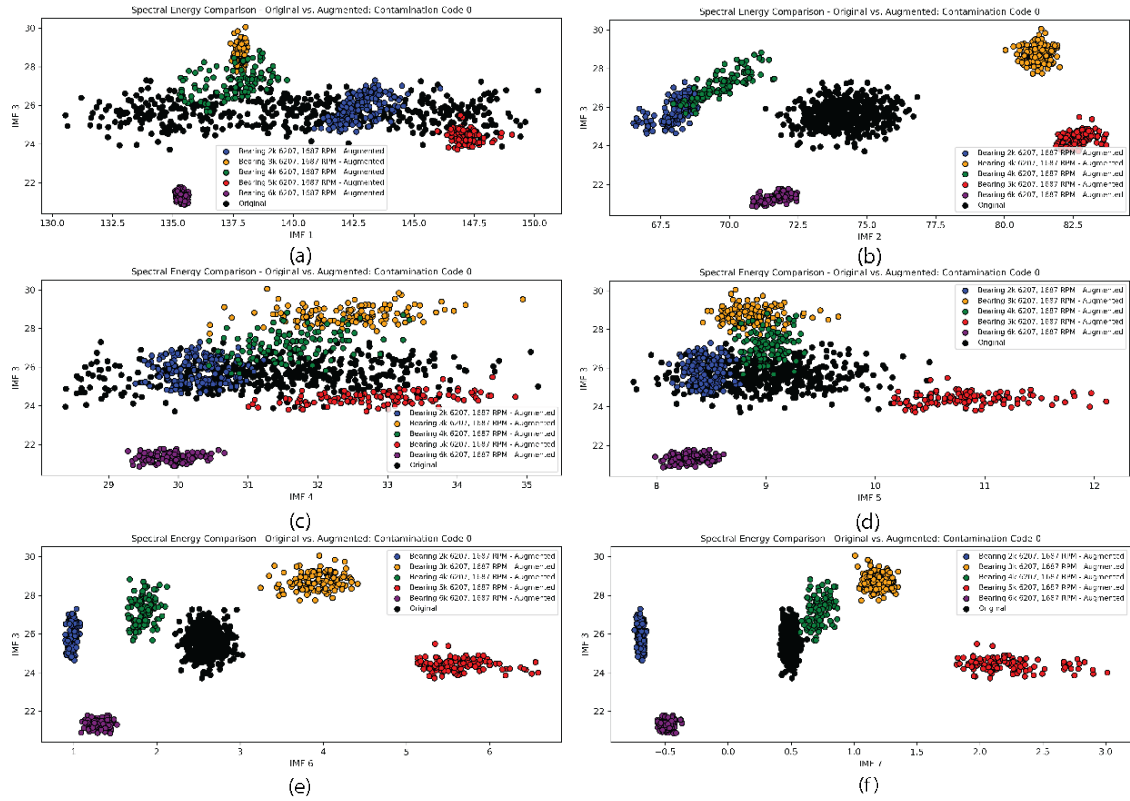
Figure 5.48: IMFs for the augmented bearing data for each 6207 case bearing versus the original bearing data in the contamination code 0. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

increased in the system as it occurred with the contamination spectral deltas. However, there does appear to be a more noticeable amount of frequencies in the higher frequency range greater than 1000 Hz, which may provide a clearer indication of a progressing fault. The primary purpose for this section is to verify and ensure that the spectral augmentation creates representative fatigue failure data from the spectral difference between the average fatigue and baseline data.

Figure 5.50 shows the spectral augmented bearing data for the 1 mm defect case compared to the true bearing data collected from the experimental testing. The baseline variance retains a similar shape as to what was observed in the baseline bearing data. What is interesting is that for the fatigue case, the data shape appears to match closer to the individual bearing baseline cases based on the potential variation seen in the bearing data for that particular frequency range. That observation may stem from having a similar frequency profile that just increases with the relative damage. However, that does not mean that the frequency peak locations are the same across the
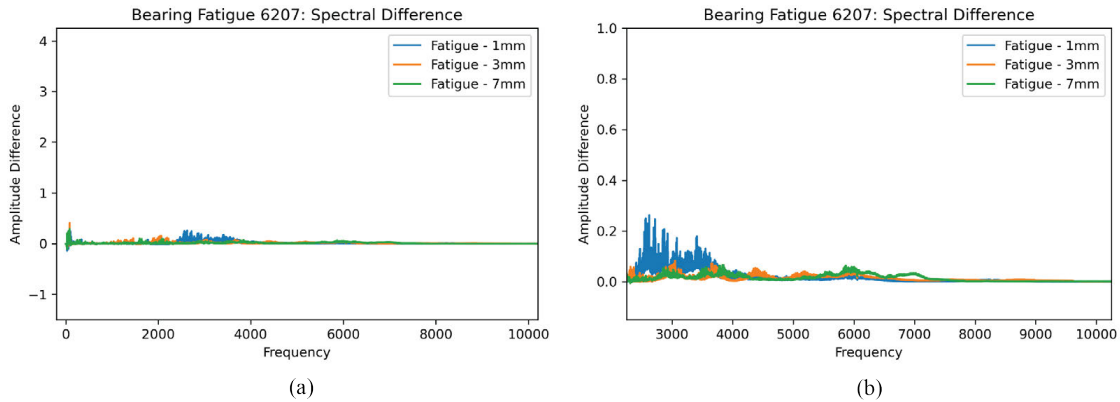
Figure 5.49: 6207 fatigue (a) spectral difference for the full spectrum and (b) zoomed into the higher frequency ranges

different cases. For the IMF energy values, the IMF 1, 2, and 5 cases appear as larger than the baseline instance. The IMF 1 and 2 appear to signify the early inception of damage that occurred with the 6207 contamination data in the Code 20 case. For the IMF 5 case, an increase could stem from the rolling elements impacting the defect as that particular range contains the bearing characteristics frequencies. However, it is interesting to note that the bearing defects do not appear to increase along the frequency range as expected.

Figure 5.51 shows the spectral augmented bearing data for the 3 mm defect case compared to the true bearing data from the experimental testing at the 3 mm defect case. For this case, it was expected to see a larger impact on the bearing characteristic frequencies (IMF 5). However, in this instance, there is a continuing increase in the IMF 1, 2, 3, and 4 range relative to the baseline case. The increase in the higher frequency range is expected as the rolling element rolling over the defect should cause additional spikes in the data. Hence, the reason for the increase in the IMF 1, 2, 3, and 4 range. It should be noted that the increase in these higher frequency ranges should stem from the impulse events that occurred in the bearing characteristic frequency ranges. It should be noted that this is a drawback of the spectral augmentation method as it does not correct the perceived signal defects in the data. These signal defects require additional signal processing to alleviate the expected signal deviations. However, despite that difference, the bearing augmented data does appear to intermix well with the IMF 1, IMF 4, and IMF 5 cases. Consistently, the augmented data from Bearing 5k appears to not intermix as well with the data range. This occurred in the 1 mm case and the contamination data. The indication of this variation could signify additional noise that

290
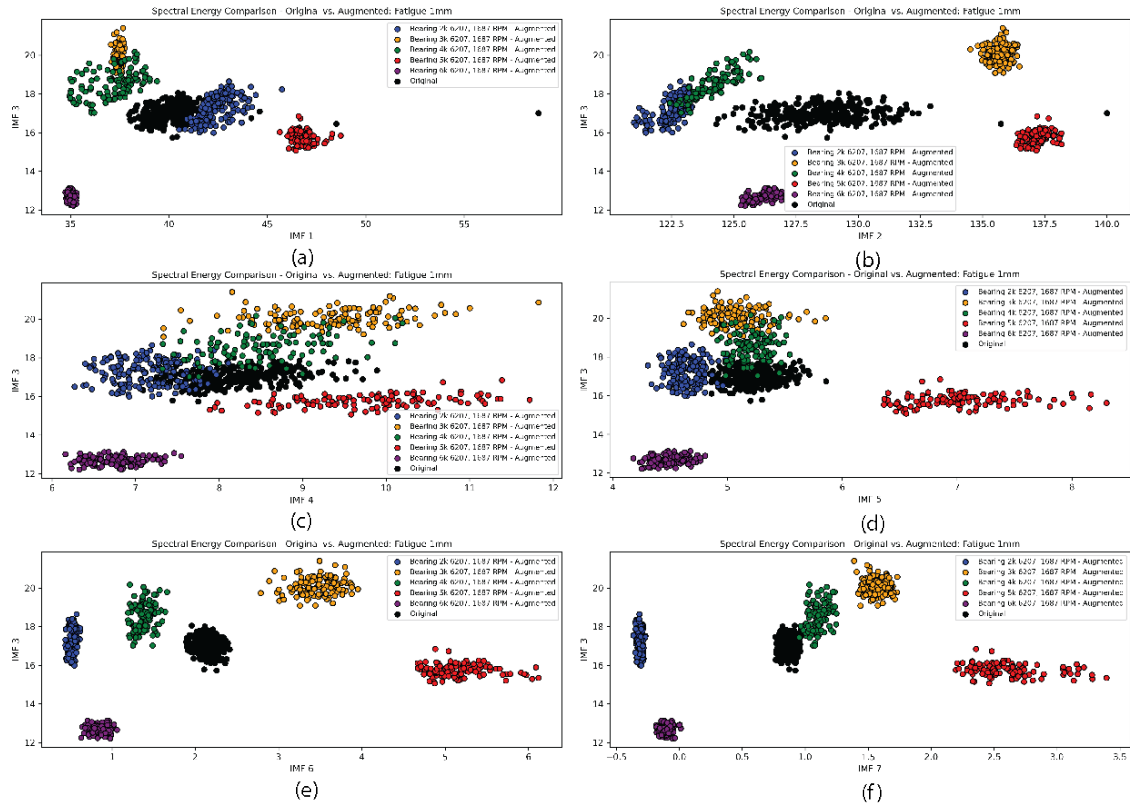
Figure 5.50: 6207 fatigue case IMF energy levels for the spectral augmented bearing data versus the original bearing data at the 1mm defect. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

is affecting this data over the other augmented cases. Other notable instances of variations occur with Bearing 3K in IMF 2 and Bearing 2K and 6K in IMF 6.

Figure 5.52 shows the spectral augmented bearing data for the 7 mm defect case compared to the true bearing data from the experimental testing. From the 7 mm case, the bearing IMF energy ranges increased for the IMF 1, 5, 6, and 7 cases relative to the baseline case. The largest increase appeared to come from the IMF 1 case, which is similar to the 1 mm defect case. Returning to the spectral failure deltas for the fatigue case (Figure 5.49), IMF 6 and 7 appear to increase as a result form additional impacts in the bearing frequency range. At this defect stage, a bearing rolling element is in contact with the defect at least 90% of the time based on the physics modeling from Section 5.1.4. Relative to the baseline case, the IMF energy 1, 2, 3, 5 and 6 are greater than the baseline case. Hence, the only clear progression for the spectral augmentation is the IMF 1 energy band for the failure fatigue range. Similar to the other defect stages, the damage cases appear to

291
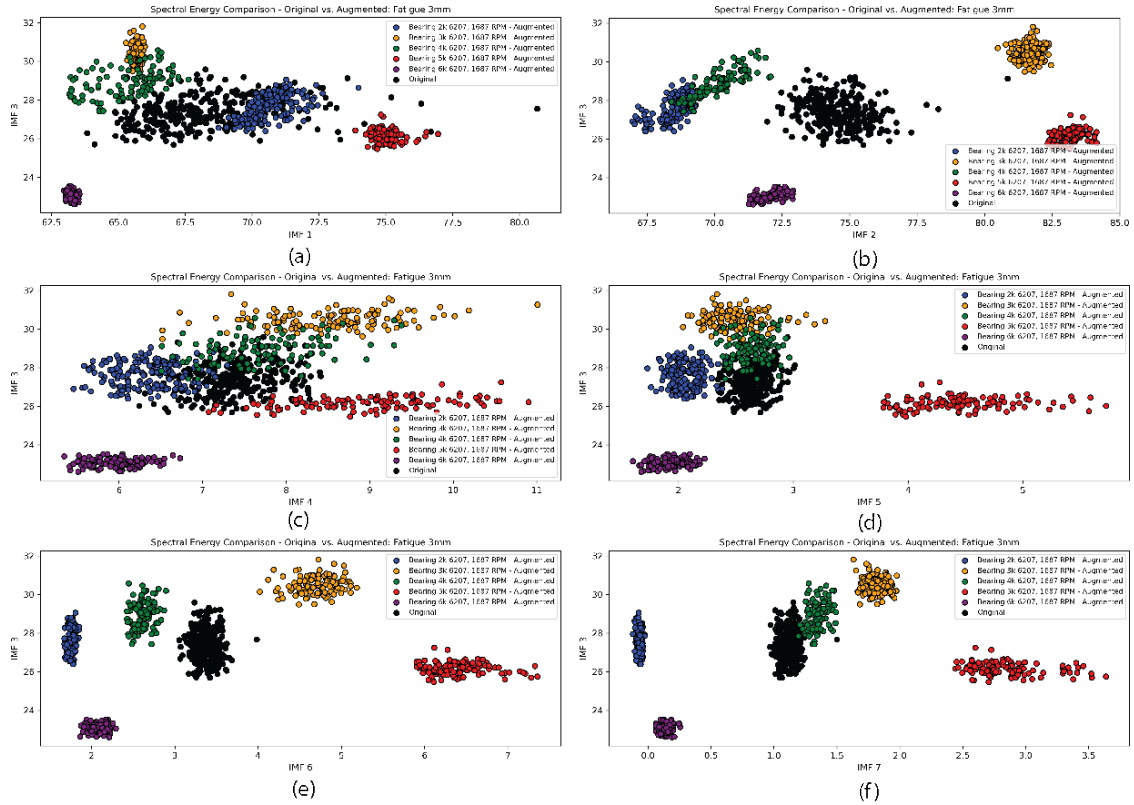
Figure 5.51: 6207 fatigue case IMF energy levels for the spectral augmented bearing data versus the original bearing data at the 3mm defect. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

coalesce around the real defect stage cases. The approximation overall for the fatigue case does appear greater in the variation as opposed to the contamination as a result of the constant damage size. For the contamination, the variation in damage appears to change as a result of the length of time associated with the test.

5.2.4.3   Fatigue: 6206

Additionally, the 6206 fatigue bearing data was used to compare the spectral augmented versus the original data. The 6206 data shows the effects of spectral augmentation with a different bearing size over the fatigue failure mode. The formulation for the average FFT and the spectral difference was conducted similarly. Hence, an average baseline case was created for the 6206 case, similar to the 6207 case. As shown in the subsequent figures (Figure 5.54 - 5.56), the bearing data does appear to have similar shapes between the different baseline cases. Figure E.1 shows the 6206
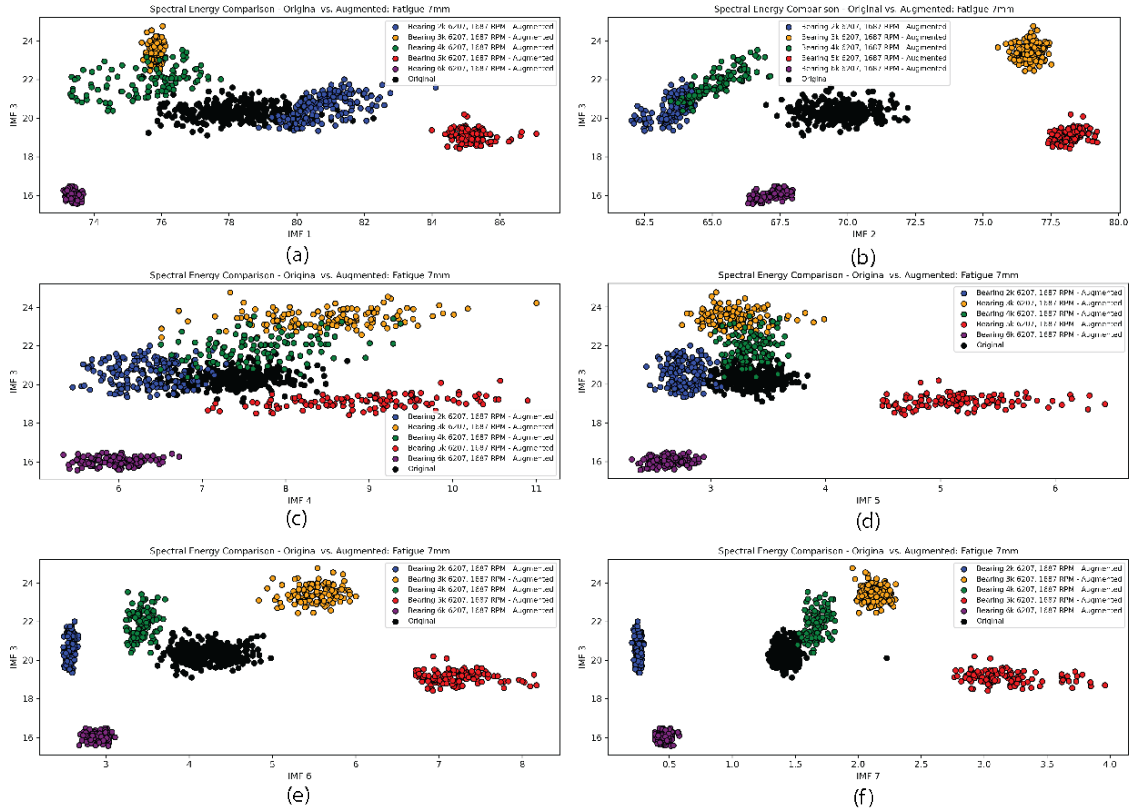
Figure 5.52: 6207 fatigue case IMF energy levels for the spectral augmented bearing data versus the original bearing data at the 7mm defect. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

representation for the baseline case based on the different IMF cases.

Figure 5.53 shows the average FFT and the spectral failure delta for the 6206 case based on the different fatigue cases. In the average FFT case, the bearing appears to first appear in the high-frequency case, similar to the 6207 case, before rising rapidly in the low-frequency case. The bearing data then increases quickly in a middle-frequency case at 2500 Hz, in addition to the low-frequency case. The spectral differences shown in Figure 5.53 (b) show that the low frequencies still have amplitudes less than the baseline case. The differences change as the length of the bearing damage increases. In the higher frequencies ($\gtrsim$ 1000 Hz), the bearing spectral differences increase as the damage increases, similar to the contamination case. It does not reach the expected plateau observed with the contamination data.

Figure 5.54 considers the different IMF energy levels for the 1 mm defect case at the 6206 case for the 1687 RPM speed case. For the IMF energy levels, the IMF 1 case is the only case that
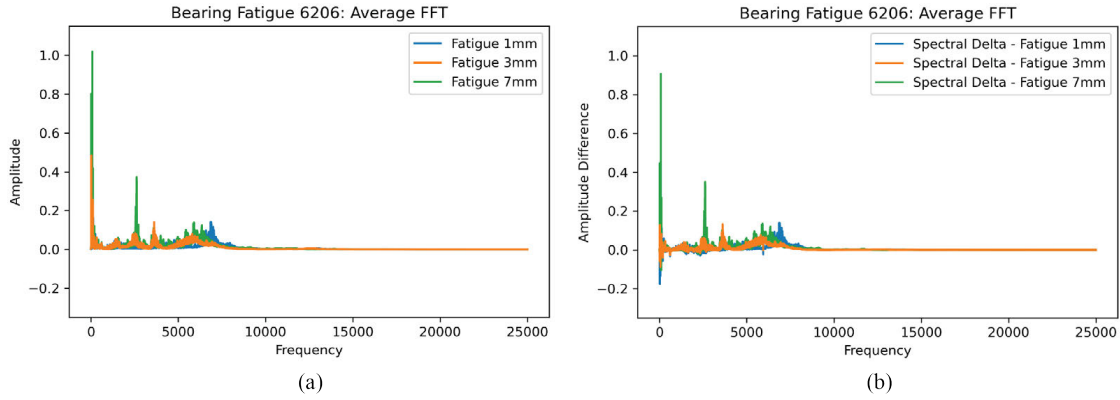
Figure 5.53: 6206 fatigue (a) average FFT and (b) spectral failure delta for the fatigue case

increases substantially relative to the 6206 baseline case in the case for the true data. The bearing baseline cases appear to vary substantially around the 6206 case. The variation appears to stem from the wide variance in the baseline cases. The individual augmented cases do appear to retain a similar shape as was seen in the 6207 case for every case except the IMF 1 case. That could signify a difference in the frequency content relative to the defect case. Outside of the IMF 1 case, there are slight increases for the IMF 2. The IMF 3, 4, and 5 cases equal the 6206 baseline case. The IMF 6 and 7 cases are slightly less than the baseline cases. As mentioned earlier, the initial bearing case changes the higher frequency range. The decrease in the low frequencies for the defect stage is anomalous and could stem from noise in the test stand. However, when using the spectral augmented method, it is possible to still retain similar characteristics with the bearing data. Hence, noise is transferred in instances when the original datasets are considered noisy.

Figure 5.55 considers the different IMF energy levels for the 3 mm defect case between the real and augmented features at the 1687 RPM speed case. In the 3 mm case, there is no difference in the IMF 1 energy level in the 1mm case. There is a lower amount of variation between the different augmented cases. For the IMF 2 case, there is a large variation that occurs between the different cases. In the IMF 6 and 7 case, the augmented feature cases do appear to retain the shape of the real features for the different energy cases. A similar shape is noted in the IMF 4 case as well. There is an increase in the IMF 2, 3, 6, and 7 cases. The increase could be tied to the increase in the defect length. However, the IMF 4 and 5 cases for the bearing fatigue defect. The lack of change could be due to the difference in critical frequencies in the lower and higher frequency ranges. It is possible though that the gradual increase in the defect could lead to a higher chance of detection as

294
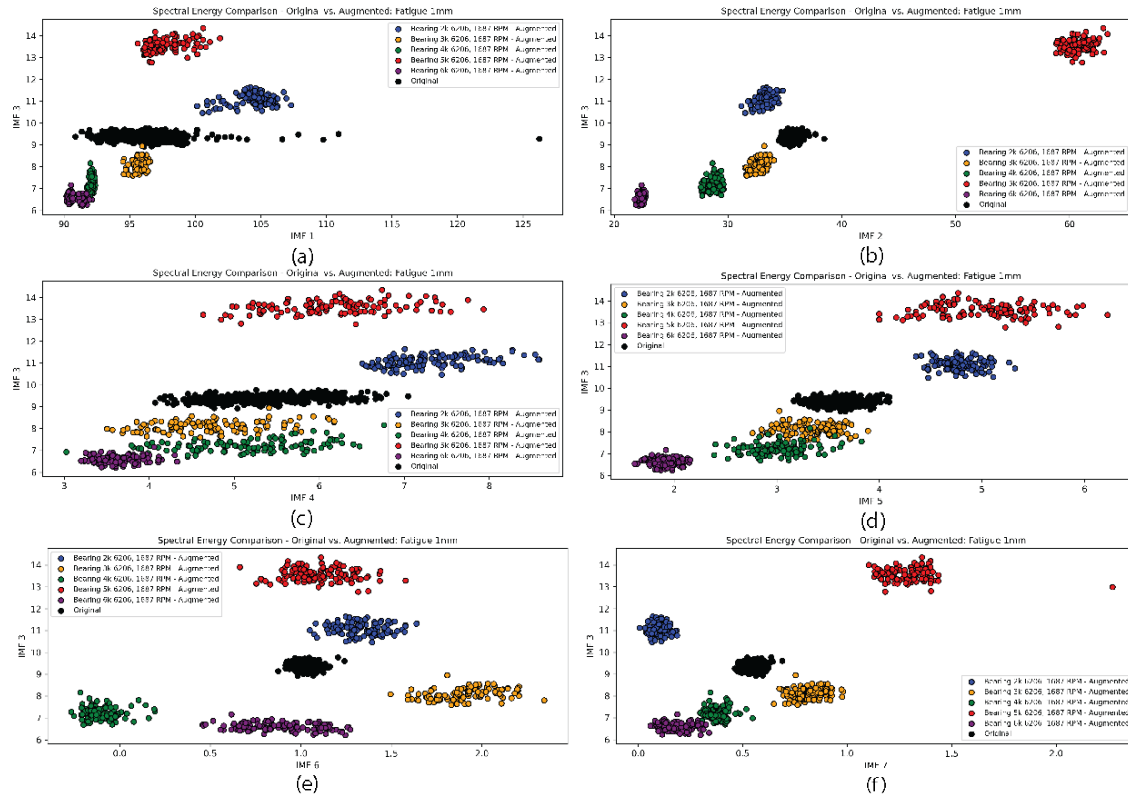
Figure 5.54: 6206 fatigue case IMF energy levels for the spectral augmented bearing data versus the original bearing data at the 1mm defect. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

the bearing defect grows.

Figure 5.56 considers the different IMF energy levels for the 7mm defect case at the 1687 RPM speed case. In this instance, the IMF 1, 2, 6, and 7 case increases relative to the 3mm case. IMF cases 3, 4, and 5 for the 7 mm cases remain similar to the 3 mm cases. The subtle increase in damage for the lower frequencies could also stem from an increase in noise from the system due to the increased bearing defect size. A steady occurrence is the additional noise in the middle frequencies found within IMF 3. The increase in frequencies within IMF 3 could stem from additional harmonics from the bearing defect frequencies in IMF 6 and 7. The overall data shape does appear to relate between the real and true features. The primary difference in the data does appear to stem in the IMF 1, which does not retain a similar shape as the real data cases.

The primary difference between the contamination and fatigue cases is the difference in the sub-clusters for each IMF. In the contamination case, three sub-clusters were found due to
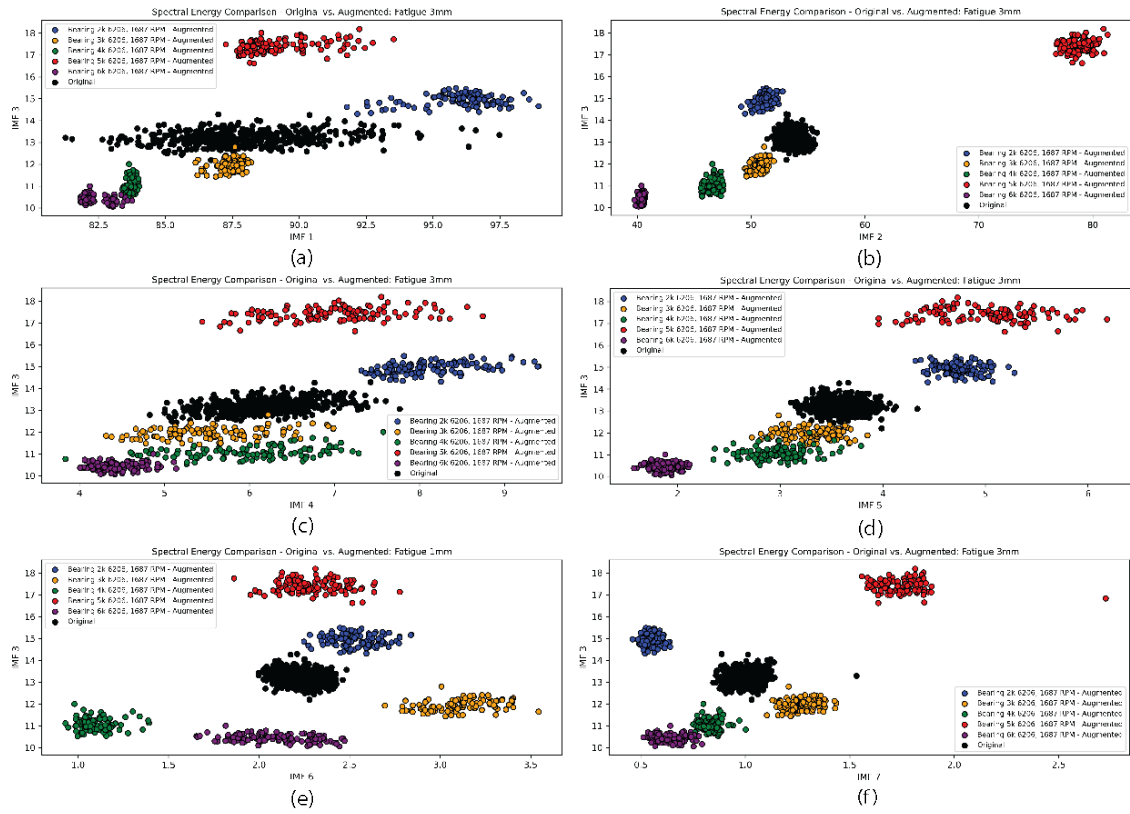
Figure 5.55: 6206 fatigue case IMF energy levels for the spectral augmented bearing data versus the original bearing data at the 3mm defect. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

the grinding of the particles that changed the different vibration signatures. In the fatigue data case for the 6206 and 6207 cases, there was one comprehensive cluster due to the testing. This observation clearly indicates that the damage conditions do not change as the bearing fatigue case is run and the more likely reason for changes in the contamination case are the result of contamination particles. In addition, whereas the contamination case high frequency (IMF 1 and 2 cases) are decreased after removing the contamination particles, the fatigue case damage remains, indicating the possibility of additional indentations from the bearing surface rather than any contamination. Similar to the contamination case, though, the augmented bearing data intermix with the original data for the fatigue case. The primary difference, though is that the fatigue case augmented features do not appear to approximate the true bearing case due to a lower variation in between the different collections as was noted in the contamination case.

The cases shown above are meant to serve as a representation of how the bearing data are

296
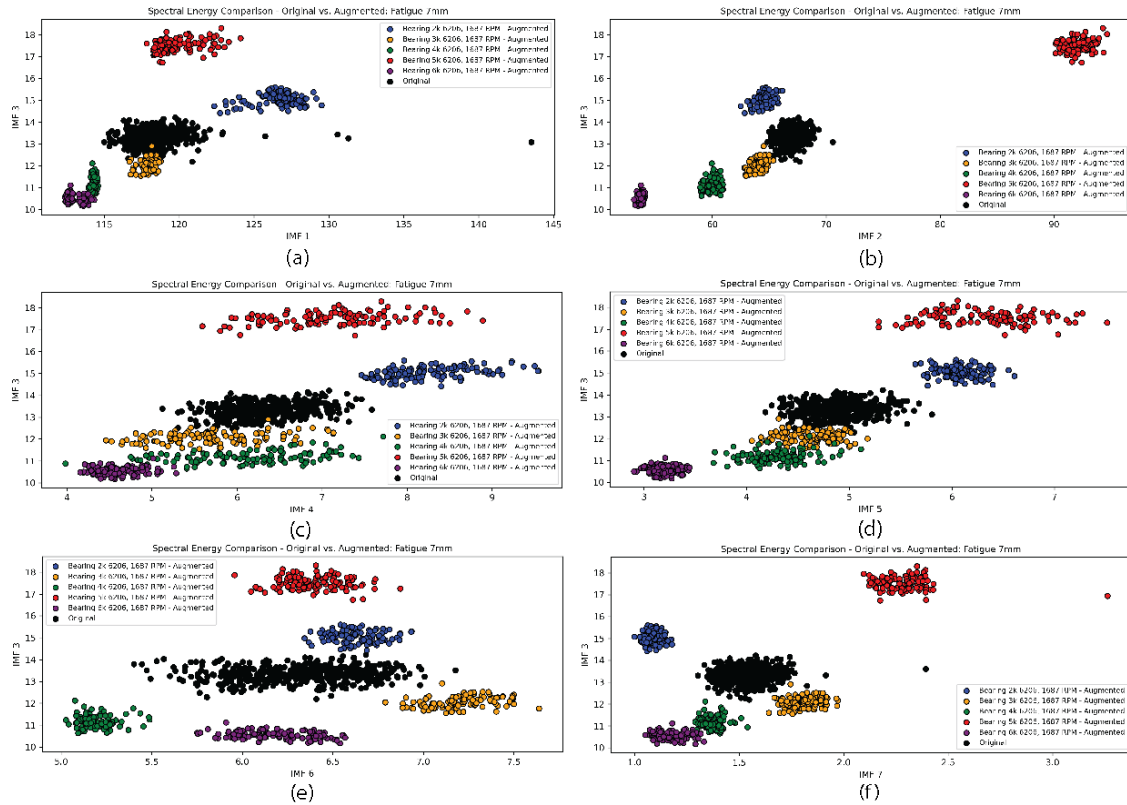
Figure 5.56: 6206 fatigue case IMF energy levels for the spectral augmented bearing data versus the original bearing data at the 7mm defect. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

manipulated to change the bearing data from one data class to another data class. The benefit of using only the baseline data is that the data are considered a "clean slate" because no known defects or impurities have been applied to the collected data at that moment. Furthermore, by establishing the relationship using the spectral augmented bearing data from the baseline data, it is possible to remove some of the variation expected with the different defect classes. As seen with the contamination data, there are instances where variation occurs within a defect class drastically. Even in the baseline case, there is variation from the different system noises affecting the bearing data. Noise and variation is an inherent part of signal processing and by starting with the cleanest dataset (*i.e.,* the healthy baseline data), there is a lower chance of introducing other variations with the dataset.

5.2.4.4 Similarity Scores between the Spectral Augmented and True Failure Data: Kruskal Wallis and Levene Test

Table 5.19 and 5.20 contain the p-values for the Kruskal Wallis test and Levene test, respectively, for testing the similarity between the true and augmented data. For each statistical test, the individual frequency bands were separated as a comparison between the real and augmented features for each failure case. Contamination Code 0 had the most instances of where the bearing data appeared to have a similar distributions at 4. Contamination Code 20 had 3, Code 24 had 2, and Code 25 had 2. For the Fatigue Case, the 1 mm defect case had three instances of similarity, whereas the 3 mm and 7 mm defect cases only had two instances. In this instance, it is not enough to state that the overall frequency feature vector is statistically the same. However, individual IMFs across the different fatigue cases do have statistically significant similarity between the real and augmented cases.

Table 5.19: Kruskal Wallis test p-values for each IMF feature. Comparison of true and augmented data for the 6205 bearing failure cases

Failure case comparison	IMF Band 1	IMF Band 2	IMF Band 3	IMF Band 4	IMF Band 5	IMF Band 6	IMF Band 7
Cont. C20	0.000	0.316	0.656	0.034	0.078	0.003	0.000
Cont. C24	0.188	0.000	0.000	0.000	0.845	0.000	0.018
Cont. C25	0.000	0.005	0.013	0.019	0.185	0.000	0.012
Cont. C0	0.000	0.000	0.600	0.148	0.283	0.965	0.709
Fati. I.R. 1mm	0.000	0.000	0.000	0.016	0.098	0.800	0.420
Fati. I.R. 3mm	0.000	0.877	0.000	0.233	0.046	0.000	0.023
Fati. I.R. 7mm	0.000	0.609	0.000	0.021	0.235	0.000	0.024

In Table 5.20, the bearing data does not perform as well in the consideration of Levene's test. Variance in forms how well the data performs in regard to the spread with respect to the average. Hence, in this test, there is not a statistically significant similarity between the variance of the real versus the augmented features. Hence, the dispersion of values is not similar between the different data cases. Similar to the IMF range, there are certain instances where there is a degree of similarity between the different real and augmented case. These particular IMF ranges may provide

Table 5.20: Levene test p-values for each IMF feature. Comparison of true and augmented data for the 6205 bearing failure cases

Failure case comparison	IMF Band 1	IMF Band 2	IMF Band 3	IMF Band 4	IMF Band 5	IMF Band 6	IMF Band 7
Cont. C20	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cont. C24	0.595	0.000	0.000	0.000	0.000	0.000	0.000
Cont. C25	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cont. C0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fati. I.R. 1mm	0.012	0.000	0.000	0.000	0.001	0.192	0.000
Fati. I.R. 3mm	0.000	0.000	0.000	0.000	0.000	0.861	0.000
Fati. I.R. 7mm	0.000	0.000	0.186	0.000	0.000	0.000	0.000

5.2.4.5 Similarity Scores between the Spectral Augmented and True Failure Data: Silhouette Index

The fatigue and contamination classes were assessed using the silhouettes score to see how well the clusters intermixed with one another. As a reminder, the silhouette score assesses the goodness of the cluster and how well they are distinguished. The score scales from -1 to 1, where -1 means that the cluster labels are flipped, 0 means the clusters are not distinguishable, and 1 means the clusters are completely separate. Table 5.21 contains the silhouette scores with the complete cases of the augmented data are compared against the available real training data. The most variation appears to stem in the fatigue case as compared to the contamination case

Table 5.21: Silhoutte Score for the final values

Bearing Defect Case	6205	6206	6207
I.R. 1 mm	0.513	0.694	0.702
I.R. 3 mm	0.763	0.655	0.677
I.R. 7 mm	0.784	0.738	0.736
Code 20	0.429	0.031	0.017
Code 24	0.113	0.051	0.119
Code 25	0.345	0.083	0.059
Code 0	0.205	0.389	0.338

The silhouette score is influenced by assessing how each feature fits within each class and is determined based on the class distinction from the furthest point relative to the center. Hence, when

there is a large amount of variance in either the true or augmented data cases, a better silhouette score (closer to 0) is possible given the possible overlap between the different classes. For example, the contamination cases, that exhibit silhouette scores closer to 0 coincidentally were noted to have a larger possible variance as a result of the different sub clusters formed during testing. In contrast, the bearing fatigue cases had only one cluster and a much lower variance on a case to case basis. Hence, there was a lower amount of overlap between the two cases as a result. This observation is further exacerbated by the variance in the baseline data. It is possible that by normalizing the bearing cases to the average initially a better result would be formed for the inner race cases.

5.2.5   Spectral Augmentation Discussion and Limitations

The spectral augmentation shows a process for augmenting bearing data using the frequency representation between the different failure cases versus the traditional method for the time series data. The method is considered a physics-informed process as it relies on the bearing system's inherent geometric and kinematic properties to modify the bearing data to create new bearing data. From the figures provided and the analysis of the silhouette scores, there is evidence that the bearing data generated and created intermix well between the two cases. Incidentally, the modeling method removes the need for physics modeling if there is some representation of failure data between the different cases. It does allow for the investigation and generation of new data for the different contamination data phases.

There are a couple of areas for improvement in the method, though, from adopting the other data-driven attributes. The first is the consensus on the average FFT, a data-driven measure. In this instance, the similarity of the bearing data could be attributed to similarities in the system parameters. However, for the FFT relatability to other scenarios, the bearing data would need an understanding of the physics, as proven in the physical modeling conducted in earlier sections. An established consensus would allow for a higher similarity score between the true and augmented cases and would reduce the possibility of introduced error from the augmentation case.

Future work for the bearing spectral augmentation method involves incorporating more physics-based measures to explain and allow for approximations between the different bearing cases. One method is using the frequency ratio to allow for data scaling between the different bearing speed cases. This aspect is feasible as the bearing speed is a consistent operating parameter collected during the bearing operation. The other aspect incorporates the physics simulation model to tune

any augmented bearing data between the different classes. The incorporation of the model helps approximate the bearing amplitudes between different cases in the event there needs to be more failure data present. Combining the spectral augmentation method and aspects of the physics model could further help approximate the shift between different systems.

5.3 Physical Scaling and Spectral Augmentation Combination

Bearing vibration is heavily dependent on speed and load, which are necessary known quantities for the design of each bearing application. The speed and geometry determine the shift in the bearing frequencies for different sizes, whereas the amplitude of each peak is determined by the load distributed through the different frequency content theoretically. Hence, it is possible to relate the speed between different applications based on ratios related to the frequencies and if there is an understanding of the load present. The basis for scaling the FFT data between the different defect cases stems from the similar frequency content between different bearing sizes based on the characteristic frequency calculations from the bearing geometry.

The standardized geometry stems from industrial standards to ensure that bearings are interchangeable between manufacturers. Figure 5.57 shows the characteristic frequencies based on the different bearing sizes. Between the 6202 and 6204 bearing sizes, the inner-to-outer race ratio remains the same, where the inner race diameter is 0.42 times smaller than the outer race diameter. For the 6205 to 6208 class, the difference between the inner and outer race diameters is 0.48. For the 6209 to 6212 class, the difference is 0.52. The variation may seem small between these different bearing sizes; however, within each group, the bearing characteristics frequencies are similar. More importantly, it provides a means of extending a ratio to shift the bearing frequencies between bearings of either similar or different sizes based on the characteristic frequencies.

Figure 5.58 shows the different bearing characteristic frequencies relative to the bearing size at 1687 RPM. Similar to the 1103 RPM case, the bearing characteristic frequencies are grouped based on similar bearing sizes. It should be noted that the inner race diameter for the bearings only increases by 5 mm through these different cases. The increasing outer race diameter is the primary contributor to the different variations in size as a result. For example, between the 6205 and 6207 bearing, the outer race increases by 10 mm from 52 mm to 72 mm. However, at the 6208 bearing size
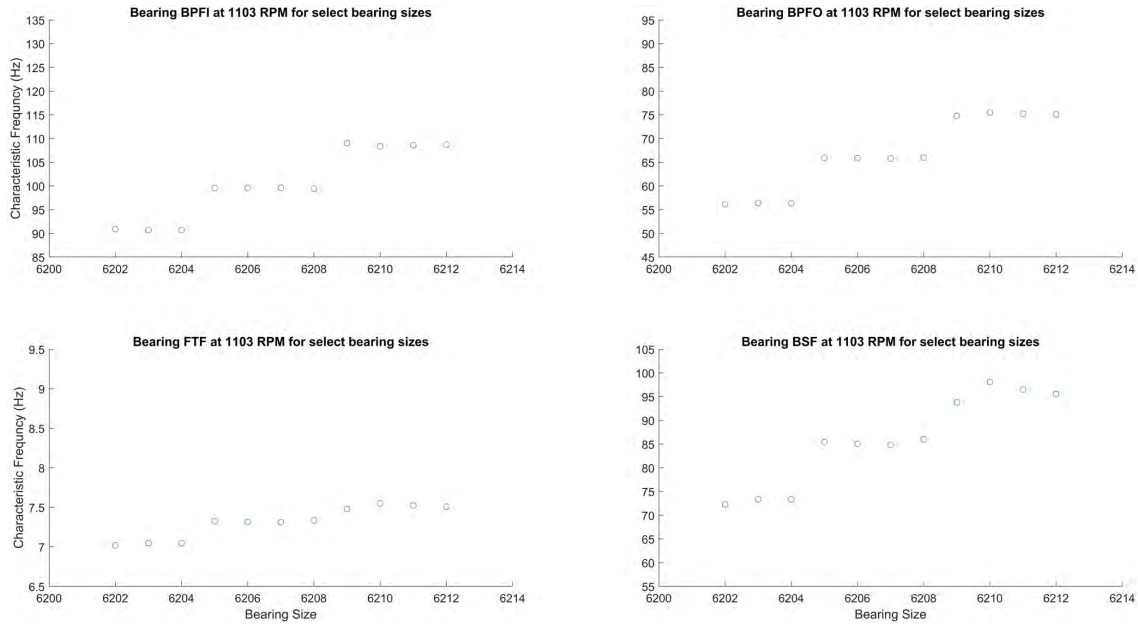
Figure 5.57: Bearing frequencies by bearing size when the speed is at 1103 RPM (18.3 Hz)

to 6211 size, the change in outer race diameter changes to 5 mm between bearing sizes. Hence, the outer race diameter and the number and size of the rolling elements are the principal components to the change in the bearing frequencies rather than the inner race, which provides a steady rate of change.

An example is provided for how the vibration frequency of a damaged bearing is shifted from a different speed condition. In this instance, data is taken from the 1103 RPM and the 1687 RPM cases. The 1103 RPM case frequency case is "stretched" along the x-axis based on a ratio determined by the speed difference. It is noted that the shift in the frequency content for this experiment considers that the FFT resolution is the same between the different defect considerations. If this is not assumed, then the frequency stretching is limited based on the particular speed case with a lower data amount and sampling rate. Otherwise, in addition to this rule, the data are limited based on the Nyquist frequency for the bearing. In keeping with the physical scaling process, the data are stretched first in the time domain using that process before having the frequency shift applied.

Figure 5.59 shows the bearing FFT responses for two augmentation comparisons for the inner race fatigue case. The first comparison considers the true 6205 data for the 1687 RPM speed case with the augmented data for the same case. The augmented data was created using the true
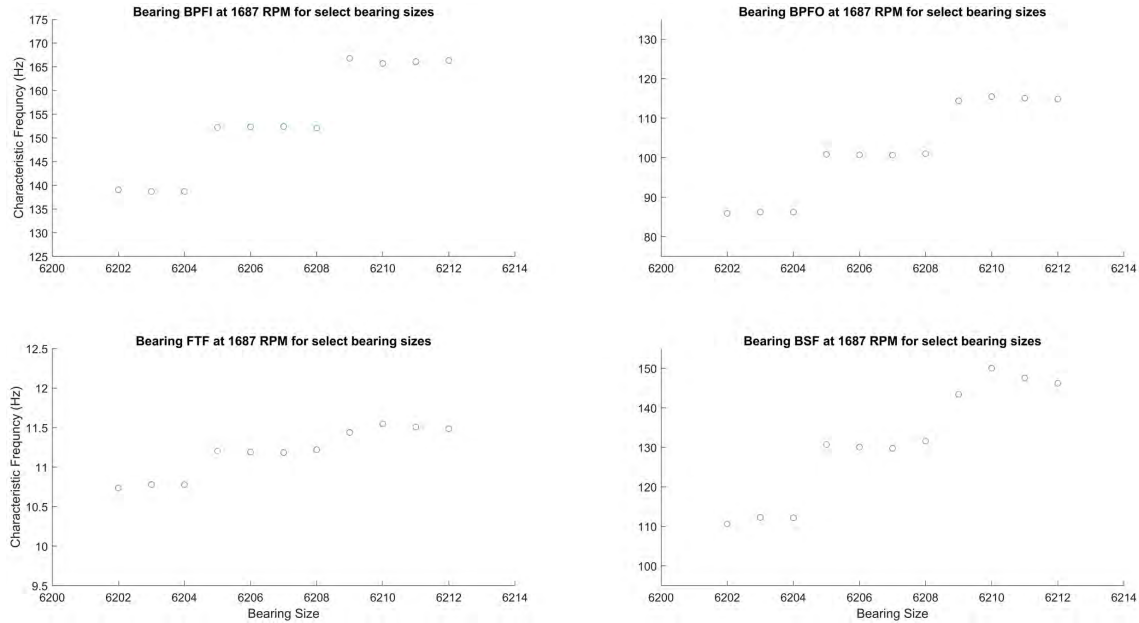
Figure 5.58: Bearing frequencies by bearing size when the speed is at 1687 RPM (28.1 Hz)

6205 1103 RPM case with the same inner race fatigue case. Based on the different data types, there are noticeable similarities in frequency positions and relative amplitudes; however, there are differences based on later frequency content. These could represent potential noises as related to the differences in noise and variance from the different tested data cases. For example, when creating bearing data based on the 1103 RPM case, there are frequency peaks that do not appear in the true bearing data that are found in the augmented bearing data. That could suggest that there are potential shifts dependent on the frequency range of the equipment.

For the comparison to the inner race fatigue case of the 6207 1687 RPM data, the frequencies appear to match even closer as compared to the 6205 case. It is possible that there is some similarity in the noise conditions experienced during the testing. However, it does demonstrate the potential merits of the idea, where the bearing data are shifted based on the frequency content and scaled based on the expected bearing response. Based on the bearing response in Figure 5.59, it is possible to modify the bearing frequency to conform to expectations related to the bearing response in the comparison of the augmented data created for the 6205 data to match the true data for the 6207 data. However, the frequency differences with the augmented data for the 6205 data to match the true data from the 6205 case, the bearing frequencies appear to hold an influence based on the noise injected from outside sources. Hence, what may be more appropriate for physics scaling and spectral
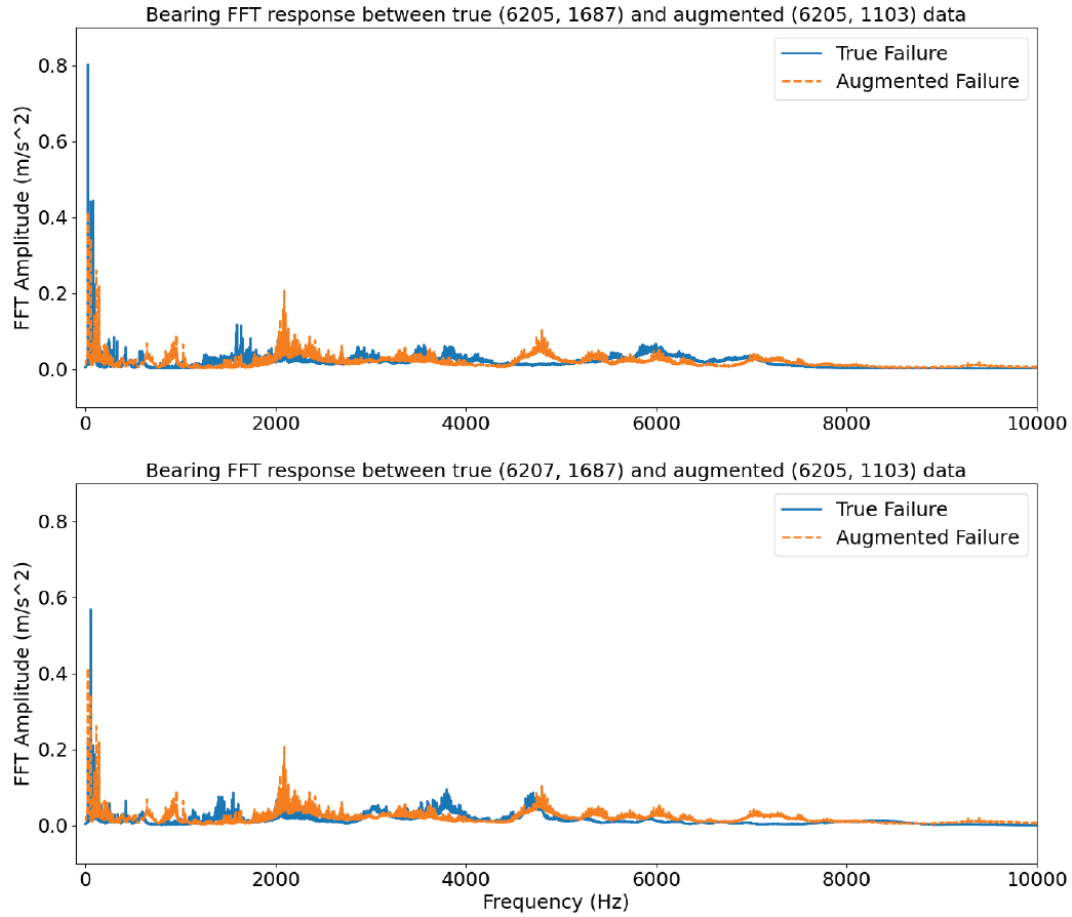
Figure 5.59: Augmented 6205, 1103 RPM data compared to the 6205, 1687 RPM data and the Augmented 6205, 1103 RPM data compared to the 6207, 1687 RPM data

augmentation is the establishment of ranges related to the likelihood of an augmented value falling within that range for the comparison of augmented and true data.

5.4 Chapter Conclusions

The overall goal was to investigate and create new methods for creating new data between the surrogate system and the production system by formulating answers to the following three questions below:

i.) Research Question 2.1: How is the artificially generated data from the surrogate system related to the production data taken from the real-world system?

ii.) Research Question 2.2: What considerations or knowledge are needed to relate different systems?

iii.) Research Question 2.3: Are the physical scaling and spectral augmentation sufficient for translating data from the laboratory environment to the real world when considering similar applications?

For Questions 2.1 and 2.2, the primary method for relating data between data from prototypes and the surrogate system is either scaling laws or dimensionless numbers from past literature. Both methods rely on physical quantities to approximate the changes between the different systems. The physical quantities are related to the system operating parameters, geometry, the dynamic properties acting on the system, and the environmental properties. One method of approximating these different aspects is using Mass-Spring-Damper (MSD) models to determine the vibration levels needed under a series of parameter considerations. Hence, as changes occur to the system, these are tracked in the MSD models and a modified vibration output is necessary. However, none of these MSD models have been used to scale data between different bearing systems or sizes.

As a result of the findings of Question 2.1, a physics model was conceived to approximate the bearing vibration response between the defect cases by verifying the bearing data response and possible corrections to increase the expected data separability in the system. The physics parameters were derived using a varying set of means to determine the expected vibration output of the system. The chosen parameters for the system stiffness and damping were much greater than those found in the literature; however, a modification factor was incorporated to prevent overshoot of the system. In truth, if that modification factor were applied to the parameters themselves, that would yield similar expectations to what was found in prior literature. However, upon application, the system did not perform as expected under the expected scenarios. In the idealized scenario, the time series data performed well, but the frequency data performed poorly (did not match the expected system response). Hence, it was attempted to force the model to have a similar response based on the experimental data. In the frequency data, the model appeared to perform as expected, whereas the time series data did not perform as well.

Due to the variation in the model, the second scenario was considered for the physical scaling process. In this process, the model output for source and target-bearing data was used as a scaling factor to manipulate either the time series response or the frequency response to produce new

and similar data. A Kruskal Wallis test was initially considered but yielded no significant results of similarity between the augmented and true data. For Levene's test, 4 of 7 features did have similarity between the true and augmented cases for the time series scaling; however, there was no greater significance in the frequency scaling.

A counterpart to the physics model approach was proposed using spectral augmentation. The bearing failure components from each defect signal were extracted and then applied to the bearing data to create augmented failure data. In the representation of augmented failure data, there is an overlap between the augmented and true failure data cases based on the Kruskal Wallis test. However, the augmented bearing data did not perform well for Levene's test. Hence, the approach serves more as an antithesis of the physical scaling process, where the opposite occurs.

Hence, for Question 2.3, both approaches show a degree of similarity between the true data and the augmented data in different methods. The physics scaling process appears to retain more of the data shape and characteristics, such as variance but appears to lack resemblance to the frequency data. For the spectral augmentation process, the bearing data matches the same distribution of the true data but loses some of the data shape and dispersion. However, both methods do demonstrate a degree of similarity is captured by predicting new failure data based on "known" failure data from a different source.

Hence, at this point, the following research questions could be considered answered based on the coloring. Table 5.22 considers the current status of the research goal as of this chapter. RQ2.3 is left as a yellow answer and is uncompleted based on the current stage of work, but is further investigated in RQ3.1 to see if the augmented data can replace the true data in different algorithmic approaches.

Formulation of a Digital Triple		
RA 1: Data Generation	RA 2: Data Transferability	RA 3: Data Analysis
RQ 1.1	RQ 2.1	RQ 3.1
RQ 1.2	RQ 2.2	RQ 3.2
RQ 1.3	RQ 2.3	RQ 3.3

Table 5.22: Current Status of the Research Goal as of Chapter 5

Chapter 6

The Digital Triplet: System Diagnosis and Prognosis

Figure 6.1 considers the complete representation of the work conducted towards the goal of a Digital Tiplet (DTr) until this point. In Chapter 4, the primary purpose was to detail and evaluate all of the data sources needed for the DTr. The first data source is the data from the production system (another mission-oriented piece of equipment) critical to operations, also called the physical triplet. The data gathered from this type of system is described as having a mix of failure and healthy data that may or may not be adequately labeled. The knowledge of that system fuels the creation of the surrogate system, which serves as the second data source. The data generated from this system represents different failure modes that could occur in the production system. After developing and labeling the respective failure data, efforts shift to determine the relationships needed to describe the bearing data.

In this sense, a third aspect of the surrogate triplet is needed: the methods for describing the data transfer from the surrogate triplet back to the physical triplet (production system). In Chapter 5, two methods were explored with the physical and surrogate triplet data to theorize different ways to scale the defect data. The first method considered using physical models to approximate the vibration of the target system (the physical triplet) and the source system (the surrogate triplet). Coupled with applying a frequency ratio to account for speed, the bearing data from one system could change to the bearing data for a separate system. A second method was explored with the spectral

Real World Test Data
- Failure Modes (if available)
- Operating Speed
- Applied Load [1]

Generated Vibration Data
Failure Modes
• Different Defect Progressions
- Operating Speed
- Applied Load [2]

Physics Model
- Physics Approximation of Test Stand
- Spectral Augmentation [3]

Gather bearing data from a real world system with the following information.

Generate vibration data to fill in the potential holes from that real world system.

Construct a physical representation from the system based on the knowledge surrounding the system

Vibration Source

Vibration Target

Physics Model Source

Physics Model Target

Original Vibration Data

Augmented Vibration Data

Formulate a series of laws and relationships based on a ratio of the source with the target physics model to establish a metirc to relate bearing vibration data between different systems.

Apply the vibration data through the model relationships to generate new bearing data based on the defect failure modes to create scaled defect data

Low amounts of failure data can be further augmented based on the spectral augmentation method

Train the bearing diagnostic and prognostic systems system based on the original and augmented data
- Consider the data amount that is necessary to train the system
- Consider whether overlapping data can substitute between different failure modes

Figure 6.1: A layout to describe the current scope of work thus far before working on the diagnostic/prognostic system training and performance

augmentation (SA) method. The SA method considers the isolation of frequency components and applies those isolated components to data lacking to create new data that intermixes with the original data. The created data increases the potential diversity throughout the collected bearing data and increases the likely model robustness.

In this final chapter, Chapter 6 describes how to leverage the augmented and original data to create diagnostic and prognostic systems. The primary emphasis is on the effect of contamination and fatigue vibration data in training these systems. Experiments were conducted with the bearing data to see how a set of classifiers and regression models operated under the separated data. Five machine learning techniques are used to test how well the data predict bearing failure: Random Forest, Decision Trees, Naive Bayes, Nearest Neighbors, and Support Vector Machine. In addition to these machine learning techniques, both Shallow, Deep, and Long Short Term Memory (LSTM) neural networks were used. Additionally, state models, such as Hidden Markov Models, are considered for the regression analysis.

The focus of this chapter is to provide further analysis related to determining between healthy and failure data. The chapter is focused on the following questions:

i.) Research Question 3.1: Is there an ideal amount of real versus augmented failure data for deploying offline data to a bearing diagnostic system?

ii.) Research Question 3.2: What augmented data process (data augmentation or physics scaling) increases the model performance regarding the assigned algorithm metrics?

iii.) Research Question 3.3: Does feature window optimization for bearing features improve bearing fault diagnosis regarding the assigned algorithm metrics?

To answer these questions, a series of tests are run with the bearing data to understand the limitations related to the data performance. The classifiers listed above are split into classifier and regression models. Every model, except the LSTM and HMM, is used to classify the different bearing defect stages. In the regression case, these models are incorporated into the training process and tested alongside the other model characteristics. The experiments conducted with the data are listed as follows:

i.) Original data test to establish a baseline understanding of model performance when only considering the original data.

ii.) Assessing model performance on transferring the models to other bearing sizes and a production example.

iii.) Test models trained based on the physics model scaled data and the augmented bearing data.

iv.) Assess model performance on bearing data with smaller feature windows.

6.1 Model Descriptions

6.1.1 Classifier Model Descriptions

Bearing diagnostics, in the scope of this research, is defined as determining whether the damage is present in the bearing versus whether the damage is not present in the bearing. These two states are labeled as "healthy" and "damaged" states. Over the past thirty years of bearing research, multiple machine learning, deep learning, physics-based methods, and their combinations have been utilized to detect and differentiate the difference between bearing health states. Five different algorithms are considered in the damage progression: Support Vector Machines (SVM) ([43], [90], [215], [223], [528]), Naive Bayes (NB) ([297], [310], [319]), k-Nearest Neighbors (k-NN) ([215], [304], [310]), Decision Trees (DT) ([320], [425], [72], [204], [273], [320], [425]), and Random Forest (RF) ([529], [170], [269], [320]), and Artificial Neural Networks ([530], [531], [532]). These six algorithms represent a subset of the possible algorithms used in bearing diagnostics from machine learning techniques from Table 2.4. In addition to these algorithms, two deep learning algorithms were considered as a result of this work: Deep Artificial Neural Networks (DANN) and Long Short Term Memory (LSTM) networks. Each of these algorithms are tested based on their ability to separate out bearing healthy data versus damage data across the different failure modes. Each algorithm operates based on different principles and a short description is provided below.

6.1.2 Support Vector Machines

Support Vector Machines work by considering a data point as an x-dimensional vector (usually a feature vector) classified based on a hyperplane. The dimensionality of the hyperplane is considered x-1. Hyperplanes are constantly drawn or fitted to the data by assessing the margin between the hyperplane and different data classes. The number of hyperplanes depends on the number of data classes, where there is one less hyperplane to describe the different data classes. The

hyperplane chosen is the one with the maximum distance between the nearest data point to the line on either side. In this research, the algorithm is tuned by using the following parameters: kernel (linear, rbf) and regularization (0.1,100). These parameters are selected based on recommendations from the user guide from scikit-learn [533].

6.1.3   Naive Bayes

Naive Bayes is a probabilistic classifier that uses Bayes theory to approximate the class of a data point. As a recap, Bayes theory determines the probability of an event based on the prior knowledge fo the event. Hence, in the algorithm approach, p, the number of probabilities, is assigned to k, the number of class outcomes, that a feature vector, x, could belong to. The probability of the data point being with in a certain data class is then the probability of x being in a certain class multiplied by the probability of the class and divided by the probability of the feature vector. A Gaussian Naive Bayes from scikit-learn is used to approximate the different data classes. It is noted that the algorithm assumes a Gaussian distribution, which may skew algorithm results.

6.1.4 k-Nearest Neighbors

The k-NN algorithm determines the data class for a feature vector by using a distance metric and the class of the surrounding feature vectors to determine what class the point belongs in. Hence, the point is determined based on the closest neighbors representing a similar class. The algorithm performance is dependent based on the number of neighbors, the weighting of the neighbors, and the distance matrix used to approximate any data. The weighting influences which points, based on the distance, approximate the data better. The hyperparameters tuned in this algorithm are the number of neighbors (1, 10), the weights ("uniform", "distance"), the distance metric computation and the algorithm for calculating distance ("ball_tree","kd_tree","brute").

6.1.5   Decision Trees

Decision trees are models that are trained to determine different data classes based on simple decision rules based on the training data. Depending on the features vector and the separability between different classes, certain features differentiate classes better than others. The features that have the higher decision power usually dominate the other features and determine whether or not

they are needed for future classification. Due to the ability to create multiple decision rules, decision trees can become very long, with hundreds and thousands of rules created. The creation of these rules can cause overfitting and create a black-box algorithm, where it is difficult to determine how decisions are made. The decision trees are tuned by using the following parameters: the criterion to measure a split ("gini", "entropy", "log-loss"), the strategy for splitting at each node ("best", "random"), the maximum depth (2-2000), and the minimum number of samples required for a split point (2-200). The criterion for splitting is for partitioning the feature space that similar labels are grouped together.

6.1.6 Random Forests

The Random Forest algorithm is made up of multiple decision trees that are fit based on subsets of the data to provide an average accuracy to feature vectors. When a new feature vector is brought into the algorithm, it is split into multiple subsets that are then assigned to the different trees. Each tree has a set of decision rules that determine the class of the overall feature vector based on that subset. The final class determination occurs based on the average response across the decision trees. The primary hyperparameter added from the previous algorithm is the number of estimators (decision trees) considered (100 - 10000).

6.1.7 Neural Networks

Artificial neural networks (ANN) are machine learning techniques that process data, like how a human brain processes data through biological neural networks. An ANN is made up of nodes and connections between those different nodes. The nodes are similar to neurons that appear in the human brain. The connections between each node represent the synapses transmitting signals between the neurons. Each node in the ANN ingests the signals and then outputs a nonlinear sum based on the ingested data. The output is then passed to other neurons. The signals are continuously passed through the network until the final value or expected information is returned. Different variations of neural networks consider the number of hidden layers, the training methods, and the data manipulation. For example, an ANN is a neural network comprising only nodes and connections. However, these neural networks can be trained using both feedforward, or backpropagation. In a feedforward neural network, no information is passed back during the training process [534]. Hence,

the weights associated with each neuron are not updated if there is any error related to the system. In a backpropagation neural network, the weights are updated based on a loss function at the output layer [535]. The loss function provides feedback on how well the algorithm trained itself based on the associated weights for each hidden layer. An artificial neural network consists of the following parameters: an input layer, hidden layer, and output layer. A deep neural network considers multiple hidden layers. The multiple layers of neurons make the algorithm "deep".

Deep learning considers the use of multiple layers to provide additional feature layers. While additional layers are possible, the architecture for each layer could change. Examples of neural networks with multiple hidden layers with different architectures include convolutional neural networks [536], recurrent neural networks [536], autoencoders [536], and generative adversarial networks [537]. A convolutional neural network considers using convolutional layers to reduce the number of features at each layer. In between the convolutional layers, pooling layers are placed to condense data down into a single neuron to output the data. A recurrent neural network considers additional hidden states that pass back information through different layers. These hidden layers can be stored in gate arrays, such as Long-Short Term Memory (LSTM) models that pass the learning back under certain conditions. An autoencoder is an algorithm that decodes data to a simplified feature space, providing a simpler representation to make decisions on the original data. A generative adversarial network considers the use of generator and discriminator to create new data based on a training dataset.

For the shallow neural network, the algorithm is tuned by using the activation function ("identity","logistic","tanh","relu"). The identity activation function considers simply the identity of the original input unmodified. The logistic provides the activation function from the sigmoid. The tanh function provides the activation relative to the hyperbolic tangent function. The relu activation function is the rectified linear unit that considers what is the max between 0 and the input. The solvers considered is the standard stochastic gradient descent and the adam optimizer, which is an optimized solver. The learning rate parameters were based as either "constant", "inversescaling", or "adaptive". The inverse scaling parameter decreasedthe earning rate with each step, whereas the adaptive learning rate kept the rate constant as a long as the training losses kept decreasing. The number of nodes considered was between (10,50). For the deep learning consideration, the number of hidden layers was considered at (1,3,5). Due to the number of features being small, the maximum number of nodes and hidden layers was reduced to remain on the smaller size.

An LSTM RNN model is considered as another deep learning bearing technique. Long-Short Term Memory Networks are termed based on their ability to host thousands of timesteps within the hidden state [538]. An LSTM considers the neuron as a cell that contains an input, output, and forget gate. These gates are regulated based on time intervals that pass information into and out of the cell. Another attribute that this format solves is the vanishing gradient problem by preserving the gradient change during backpropagation. In addition to acting as a stand-alone algorithm, the LSTM can be paired with other algorithms to improve performance such as with GAN [539], Autoencoders [540], and CNN [541]. These are the two primary reasons (thousands of timesteps and the combination ability) for selecting the LSTM model over other algorithms. Due to the time-series nature of the LSTM, the algorithm is only considered for the prognostic phase.

Figure 6.2 shows the layout for a neural network. The progression of a neural network begins with the features entering the input layer, where each node (N_i) represents a feature. The data from each input node is passed to each node with an associated weight (w_i) and bias factor (b_i). For the input layer to the first hidden layer, these parameters would be labeled as w_1 and b_1, respectively. After the initial transformation, the output is typically denoted as $z_{i,j}$, i represents the hidden layer and j represents the node in the hidden layer. An activation function is applied to the data to accommodate possible non-linearity in the system. There are generally four different activation functions used with Neural Networks: the hyperbolic tangent function (tanh), the rectified linear unit (relu) function, the sigmoid function, and leaky relu. The output from that transformation is labeled as $a_{i,j}$ and used as the next layer's input. The function for each activation function is listed in Appendix F.1.

Equation 6.1 provides the generalized nodal equation for each layer of the neural network. The $f(z)$ is the activation function. The activation function is selected based on the final tuning of the model. The model outputs predictions and compares them to the training labels during the training process. A cost function is used to measure the error within the system. During the model training, the derivative of this quantity is used to approximate new weight and bias factors for each hidden layer. The backpropagation for each layer continues until it reaches the hidden layer. This generally occurs as a result of a chain rule of derivatives. The final form of the chain rule equation is provided in Equation 6.2. Combining the forward propagation as Equation 6.1 is termed, and the backpropagation in 6.2, forms the relative structure for the neural network algorithm.

314
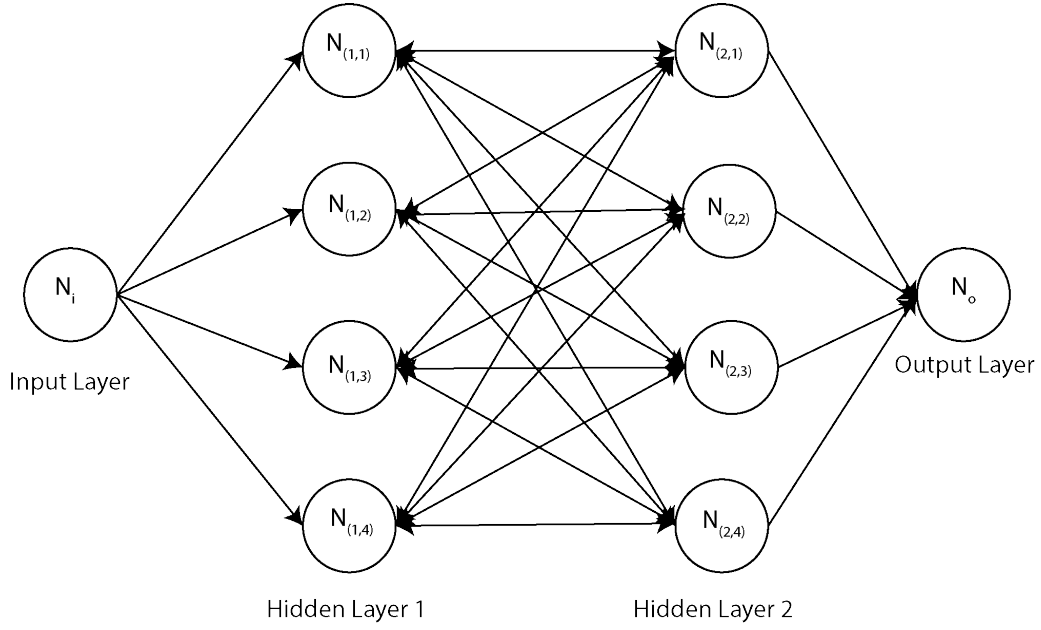
Figure 6.2: Representation for a Neural Network with a one node input layer, two hidden layers of 4 nodes, and one node output layer

$$z_{(i,j)} = w_i^t x + b$$
$$a_{(i,j} = f(z)$$
(6.1)

$$\frac{\partial C}{\partial w_{(i,j)}} = \frac{\partial C}{\partial a_i} f'(z_i) * a_i$$
(6.2)

Figure 6.3 shows the representation for one LSTM cell. There are three different inputs into the cell: $X(t)$, $h_{(t-1)}$, $c_{(t-1)}$. These are then passed to the different gates that make up the LSTM. Each gate corresponds to a different function of the LSTM cell. The forget cell determines how much information is retained from the previous data iteration. The identity cell passes all of the "useful" information from the current input. The output gate extracts useful information to pass as the output of the cell. The associated activation functions control the flow of information from the current state to the next state. The error function changes due to the different configuration as shown in Equation 6.3. The function provided considers all the different cell states and hidden states in addition to the output to conduct the backpropagation.
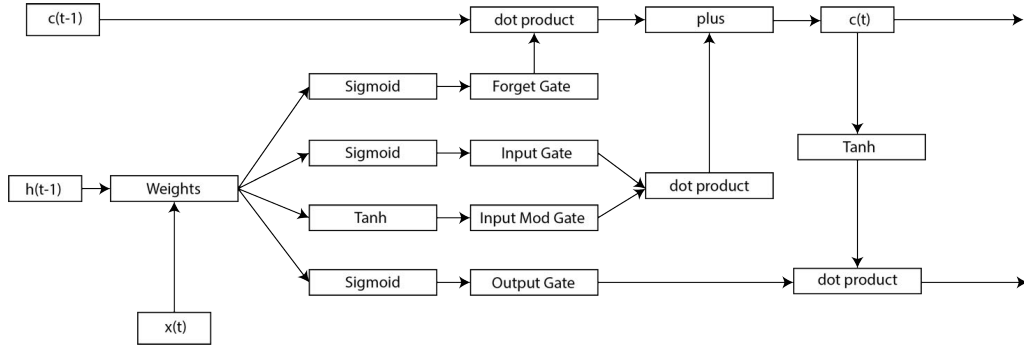
Figure 6.3: Rough representation for the LSTM

$$\frac{\partial E}{\partial W} = \Sigma_t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial c_t}{\partial c_{t-1}} \frac{\partial c_0}{\partial W} \qquad (6.3)$$

6.1.8 Regression Models

Bearing prognostics, in the scope of this research, is defined as determining the damage progression within a specific defect mode and how it relates to overall bearing health. In this essence, the different bearing stages could be characterized as a percentage. Bearing life begins at 100%. As the bearing degrades, it moves closer to 0%. Using a percentage point could be seen as arbitrary depending on the system and the vibration level range. Configuring the percentage to the expected vibration limits is possible, as ISO documentation prescribes. Physical models are possible, such as the Paris crack growth model used to describe the progression of a bearing defect on the raceway and rolling elements [344]. However, the model does not necessarily translate to conditions without a defined damage formation. Hence, the Hiddenuse of data-driven values such as percentages of damage.

This scenario considers two prognostic algorithms: Hidden Markov Model (HMM) and machine learning regression models. The machine learning regression models approximate a value versus a class state. The regression models considered are the same as the model methods considered in the diagnostic stage. In this instance, the value is a percentage of the total life remaining for the bearing based on the damage applied. A healthy bearing is to begin at 100% and then degrade over time until it reaches 0% at the final damage stage. Hence, the testable algorithms for this stage are the Support Vector Regressor, Bayesian Ridge Regressor, Decision Tree Regressor, Random Forest Regressor, and KNN Regressor, and Neural Network Regressor. The DANNR and LSTM models

316

are also considered a part of the possible regression techniques. A description of the underlying mechanics for these models were provided in the classifier section. The following subsection (6.1.9 describes the HMM method at a high level.

6.1.9   Hidden Markov Model

An HMM model consists of a Markov process with hidden states, hence the name. A Markov model/process is a stochastic model to describe a sequence of events based on the probability of a previous event [542]. A Markov model can be seen as a step over decision trees as these models can capture time-series events. For maintenance strategy [543], Markov models are employed to characterize the deterioration of equipment, where the deterioration rates are modeled as random events. An HMM contains two components, as mentioned earlier, the unobserved Markov process, U_{ms}, and a set of observable states, O_{ms}. The O_{ms} approximate the U_{ms} based on the transition and output probabilities related to the different states. Generally, the different types of condition monitoring data in bearing fault diagnosis are considered the observable states, and the bearing condition is the unobservant state. The model is trained based on similar principles of splitting data into training and test sets for building the algorithm. The hyperparameters for tuning the algorithm are the number of hidden states, the state probabilities, and the number of observant states.

It should be noted that the HMM is primarily a time-based stochastic model. The data collected in this work has associated time stamps; however, it changes between states based on the damage application versus the time conducted during testing. The condition of time series data does not discount the use of the model in identifying data clusters. Multiple instances of the algorithm are used with the Case Western Reserve University Bearing Dataset [544, 545, 546]. In these instances, the bearing data are formed into individual clusters that serve as the observed states that require determining the equipment state. A similar approach is adopted for this analysis. Liu *et al.* uses the HMM with the features extracted using variation mode decomposition, a variant of the Hilbert Huang Transform. However, other than a mention of the defect frequencies, the features are not identified. Nevertheless, a similar approach is adopted with the bearing data generated in this research.

Figure 6.4 provides an example of the HMM representation. In this instance, X represents the unobservable states, whereas Y represents the observable states. Outside of the unobservable and observable states, the B terms and A terms represent the state transition and output probabilities,

respectively. The transition probabilities are those associated with different state changes based on an initial state. A downside of this assumption is the determination that all rules considering the observable states have been included in the model during training for the proper unobservable state determination. The output probabilities are the probabilities of a certain output based on the considered hidden state. Similar to the state transition probabilities, these quantities must be included in the model for the analysis.
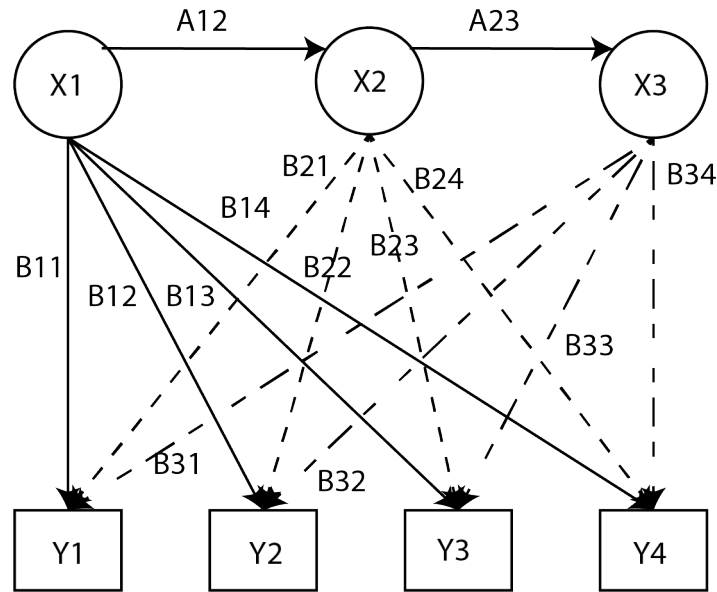
Figure 6.4: An example of the Hidden Markov Model representation

The python *hmmlearn*[12] library was used for the HMM implementation. The library considers three different model implementations. Each model considers a Gaussian implementation based on a multivariate mean and covariance matrix to assume the different probabilities. The model approximates the different states based on fitted data using unsupervised learning. Despite the assumption of Gaussian properties, including emission probabilities to better approximate changes between classes is possible. A Gaussian distribution is assumed for the initial testing unless poor model results occur from the Gaussian assumption in the classifier and regressor models for the Naive Bayes (*i.e.,* less than 75%).

Whether an algorithm is a good fit is based on the log probability of the HMM fitting to the training data. Hence, the maximum log probability score is selected due to having the best fit

---

[1]https://hmmlearn.readthedocs.io/en/latest/
[2]Software is licensed under the BSD-3-Clause

based on the samples passing through the expected observed states. In essence, the log probability of an observed state given the model is the summation of the log events given the model. Hence, the value maximized in this instance is the posterior of the real values after they pass through the model. The log probability aggregates based on the number of samples and hidden states. Hence, in this instance, it is more of a case-by-case basis between the different algorithms to see how well the model performs in approximating the different defect stages and potentially speed cases.

6.2 Feature Generation

The selection of the features used for the bearing diagnosis considers only the frequency domain due to the potential of noise from the time domain. The frequency domain features are generated based on the FFTs, where specific chunks of the FFT are split based on a frequency band split from the empirical mode decomposition. The frequency bands are divided based on the ranges as described during the Physics Scaling Process. The features are then split between training and test sets, where the data was split between 80% and 20%, respectively. The frequency energy of the vibration signal was calculated between these bands by taking the area under the FFT.

In the feature generation, the data used to create these features is changed based on the feature window size. In FFT generation, the amount of data used affects the resolution of the frequency plot. FFT resolution is determined based on the number of samples with the sensor sampling rate for each FFT. Equation 6.4 shows the calculation for the frequency resolution as the division of the sampling rate (F_s) by the number of samples (N_s). If the sampling rate is 2000 Hz, and the number of samples collected within a window is 1000, then the frequency resolution is 2 Hz/bin. The approximation also means that it is impossible to tell whether a frequency peak occurs at 2 Hz versus 2.5 Hz or even 3 Hz in this scenario. In the previous data descriptions, a resolution of 1 Hz was used. The feature window size and how it affects the frequency transform is further discussed in Section 6.7.

$$F.R. = \frac{F_s}{N_s} \tag{6.4}$$

6.3 Model Performance: Original Data

6.3.1 Baseline Data Separability and Transference

An initial test was considered to determine if diagnosing and separating the bearing data was possible based on the operating speed and the bearing size. The bearing data from the different speed cases and sizes were combined into one set and then normalized concerning the minimum and maximum values across the different IMF frequency bands. Table 6.1 contains the F1 scores for the test accuracy for the different machine learning algorithms. The F1 score is calculated based on micro-averaging, which is noted to be class biased [547]. However, in this instance, it was considered that the approximate amount of baseline data was the same across the different cases within 10%.

Table 6.1: Bearing Baseline separation utilizing machine learning: F1 score

Model	Bearing Size Diagnosis	Bearing Speed Diagnosis
Naieve Bayes (NB)	0.964	0.586
Support Vector Machine (SVM)	0.995	0.935
kNN	1.0	1.0
Decision Trees (DT)	0.991	0.998
Random Forest (RF)	0.999	0.999
Multilayer Perceptron (MLP)	1.0	0.999
Deep Multilayer Perceptron (DMLP)	0.990	0.98

The initial formulation and analysis was considered for two-fold reason. The first was to evaluate the general separability between the different classes and see if there were major differences between the different bearing sizes. The applicability of determining the different bearing sizes used in a system should be redundant, along with the speed estimation. Generally, this knowledge is available prior towards any bearing condition monitoring. However, the testing involves determining if the bearing data are discernible between the different cases.

The differentiation between the different speeds were expected based on the spread of the frequencies and increase in amplitude from the speed. The differentiation of the size was less expected. As found with the physical testing, there are similar expected levels of vibration and frequency between the different classes. However, six of the seven models (SVM, kNN, DT, RF, Neural Networks, and Deep Neural Networks) could differentiate the different bearing classes despite their normalization relative to the entire bearing dataset. The NB algorithm did have difficulty in differentiating the different classes. A possible reason is that the overlap between different features was not accounted well in the algorithm. The NB assumes that the features are independent within

each class, which is not necessarily true of real world systems.

6.3.2   Bearing Health Monitoring: Classifiers

For the bearing diagnosis considerations, a label is assigned to each class based on the defect stage and speed. Hence, there were four different data types: contamination (1), fatigue (2), brinelling (3), and lubrication (4). For the contamination stages, there were four different classes for the progression: Code 20 (1), Code 24 (2), Code 25 (3), and Code 0 (4). For the fatigue case, there are three classes for the progression: 1 mm (1), 3 mm (2), and 7 mm (3). The brinelling case has four cases: 0.5 m drop (1), 1.0 m drop (2), 1.5 m drop (3), and 2.0 m drop (4). The lubrication case had three cases: 75% lubrication filled (1), 50% lubrication filled (2), and 25% lubrication filled (3). The labels are assigned in individual progression groups. For example, the bearings damaged in the brinelling phase would have numbers to signify the particular damage phase in the brinelling label and then "0's" in all of the other label groups.

Despite having four different failure modes, only the contamination and fatigue cases are considered, initially. These cases are used as they show the most change concerning the different defect cases as they progress. The process involves first assessing how the algorithms performed on separating the different types of bearing damage. In the initial case, three different conditions could be selected: (0) for the baseline case, (1) for the contamination case, (2) for the fatigue case. Table 6.2 contains the test and train F1 score values for the different algorithms from the 10-fold grid search. Speed is not considered in this process, meaning all of the data corresponds to the defect classes regardless of the operating condition. The total data amount was split into 80% training and 20% testing. The 80% training was used in the grid search, whereas the remaining 20% was used as the validation data found in Section 6.3.2.2.

In the first instance, the model training for the NB classifier varied due to the bearing classifier. It is possible that the variance in the data occurred due to the amount of overlap that occurred between the different speed cases. It is possible that due to the probability rules the classifier could train that the data was not easily separable. This trend did not hold outside of the probabilistic considerations. The SVM case appeared to handle any potential similarities in the data. It is possible that the SVM can handle the bearing features better due to the interactions between the different features. One of the downsides of the Naive Bayes algorithm is due to the assumption of independence between the different features. However, independence between features hardly

Table 6.2: Bearing classifier performance by average F1 score for diagnosing failure type Average (Train/Test) during the grid search

Algorithm	6205	6206	6207
Naive Bayes	0.843 / 0.832	0.954 / 0.945	0.893 / 0.895
Support Vector Machine	0.999 / 0.999	0.997 / 0.997	0.999 / 0.999
k-Nearest Neighbors	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Decision Trees	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Random Forest	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Multilayer Perceptron	0.947 / 0.939	0.947 / 0.939	0.958 / 0.957
Deep Multilayer Perceptron	0.876 / 0.866	0.876 / 0.866	0.876 / 0.866

happens in real-world scenarios. Typically, failure events are related to other features within the data. The SVM assesses those interactions when selecting the different hyperplanes when maximizing the margins between different classes. The KNN algorithm is non-parametric and can also handle dependence of features, which may support why it also supports the higher classification rate over the Naive Bayes algorithm. The voting method that is the basis for the KNN can determine similar points and select the interactions based on those features.

The tree-based models also appear to support the higher performance rate. With the tree models, the classification occurs by creating a series of rules that fit the data. These rules then determine the different classes associated with the features inputted into the model. However, as a result it is possible for the features to become constrained to certain groupings and determinations where a smaller set of features determines the overall classification. Figure 6.5 shows the feature importance with respect to the mean decrease in impurity for the decision trees and random forest classifiers. For tree-based models, the mean decrease in impurity (MDI) refers to how the features are split between the different estimators, with the standard deviation shown to assess the dispersion across the range of estimators [548]. Another way to look at the figure is based on how is the overall accuracy of the model affected by the possible removal of one or more features. For example, IMF 1 in the 6206 and 6207 cases has a large associated MDI value. Removing this value may affect the overall model's ability to predict the correct class, especially in the DT classifier case for the 6207. Overall, though, for the random forest case, the feature importance is more distributed, with some high variance seen for the IMF 1 case in the 6206 and 6207 cases.

The decision tree models do not have error bars as they do not have multiple trees to assess the dispersion. However, assessing what features are based on the one tree available is still possible.
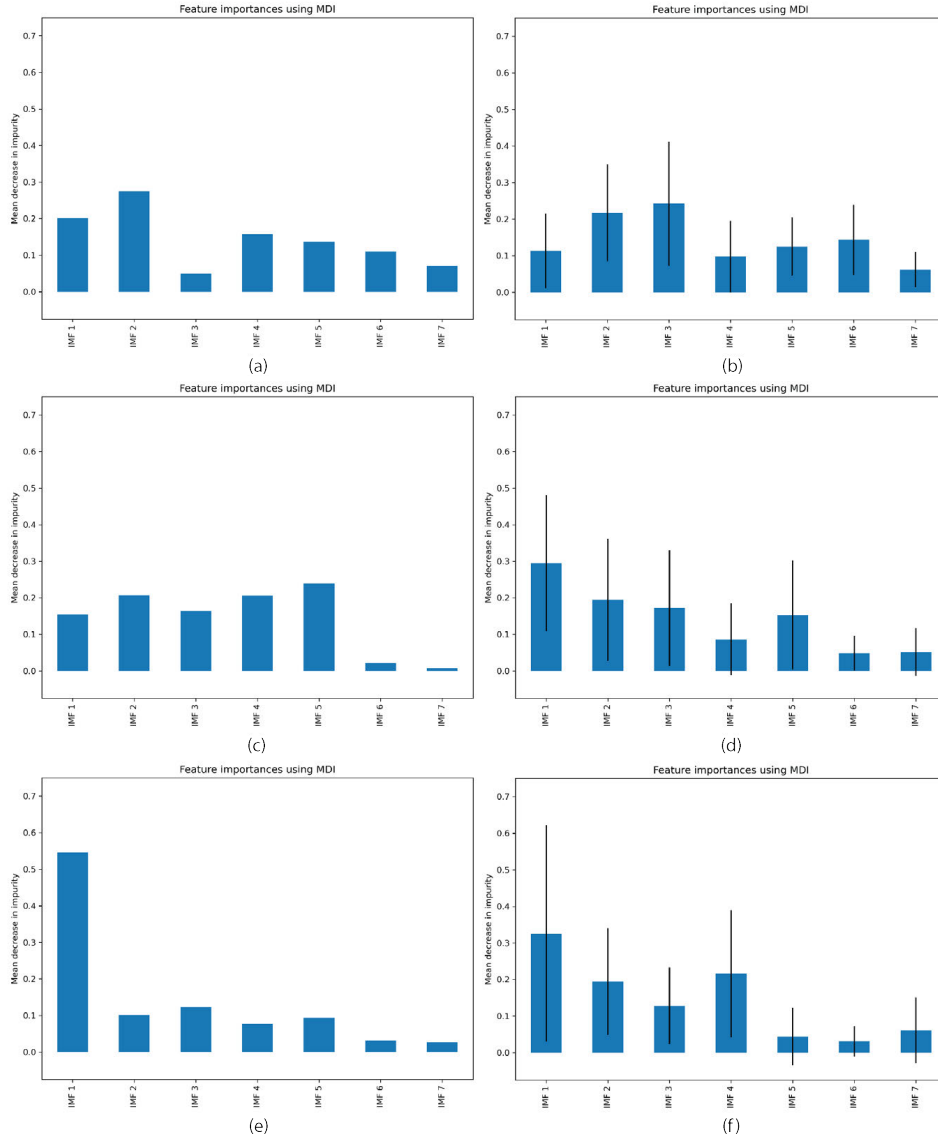
Figure 6.5: Bearing classifier feature importance for the (a) 6205 DT classifier, (b) 6205 Random Forest classifier, (c) 6206 DT classifier, (d) 6206 RF classifier, (e) 6207 DT classifier, and (f) 6207 RF classifier

For the features in each decision tree model, it is possible to see that the feature importance falls in line with what was evaluated within the range of the random forest estimators. The decision tree shows the feature importance if only one estimator was available in the random forest algorithm. In this instance, IMF 1 dominates in the 6207 case, meaning the model performance may deteriorate if that feature were removed. However, in the other bearing sizes, the feature importance appears to be more distributed. That means the removal of one feature does not appear to affect the overall

323

model performance.

Another possible method to assess feature performance is through permutation feature importance. Permutation feature importance is used when there are many unique variables, most commonly unique categorical variables. In this importance measure, the model features are shuffled to break the relationship between a target and a feature [548]. Any corresponding drop in the model score indicates how "good" or important the feature is for the model. This method was used to assess the model but yielded no meaningful results at the classification stage for the damage type.

Table 6.3 contains the average algorithm response for determining the different contamination stages. Similar to the damage types, the NB case appeared to perform similar in both cases. It is likely that the feature independence assumption is invalid in this instance as well. The 6205 and 6207 cases performed well in each of the other algorithm instances. There was a lowered instance of accuracy in the 6206 MLP models compared to the 6205 and 6207 case. That could have been the result of the difference in the approximation of speed values as compared to the different health stages. Only the tree models appear to show signs of overfitting, where the train and test scores are not close in appearance.

Table 6.3: Bearing classifier performance by average F1 score for determining contamination stages (Train/Test) during the grid search

Algorithm	6205	6206	6207
Naive Bayes	0.902 / 0.861	0.823 / 0.784	0.925 / 0.918
Support Vector Machine	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
k-Nearest Neighbors	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Decision Trees	0.999 / 0.98	0.999 / 0.999	0.999 / 0.999
Random Forest	0.999 / 0.96	0.999 / 0.999	0.999 / 0.999
Multilayer Perceptron	0.873 / 0.873	0.832 / 0.832	0.873 / 0.873
Deep Multilayer Perceptron	0.885 / 0.873	0.848 / 0.848	0.876 / 0.873

As mentioned earlier, the F1 scores show the average performance resulting from the entire algorithm performance. In some instances, the algorithm performance did vary drastically based on the parameters selected. For example, in the KNN algorithm, a parameter returned an error during the search that did not work with the data. In the SVM instance, there was no large change depending on the parameter used, indicating that it is probably due to labels in the training data. For the other bearing size cases, the only noticeable difference in algorithm performance came from the difference in kernel. The rbf kernel, which is Gaussian-based, can approximate the data classes

better, than a linear approximation.
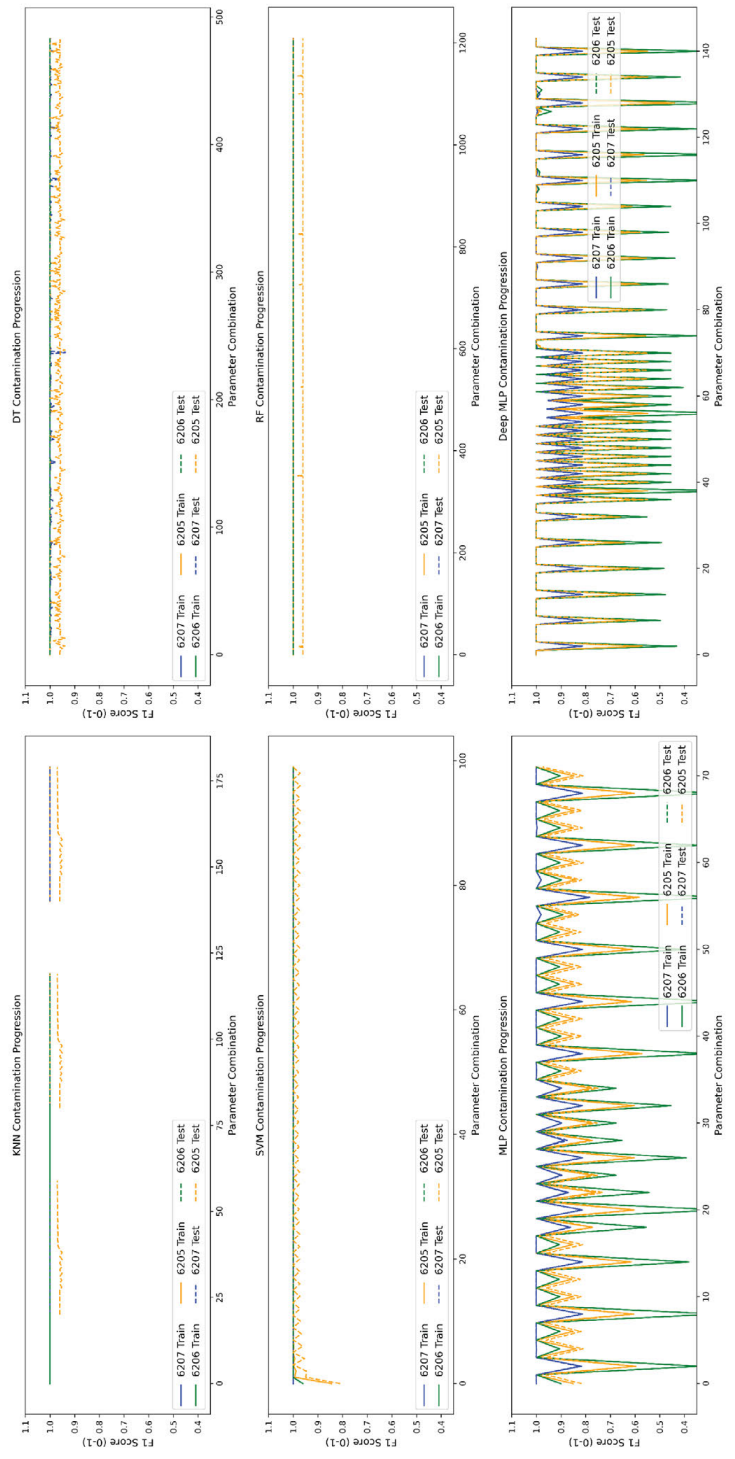
Figure 6.6: F1 training and test scores for each individually tested set of algorithm parameters used in the contamination cases

The DT and RF classifiers did experience some overfitting from the 6206 classifier. In the 6206 classifier case, the primary contributor to the overfitting and difference between the upper and lower bound was the maximum depth attributed to each estimator. However, that did not appear to be the case in the 6205 or 6207 cases. It is possible that it occurred from the By far, the largest change in algorithm performance was seen in the shallow and deep multilayer perceptron. For these different algorithms, the stochastic gradient descent algorithm for the solver was the cause of most of the poor algorithm results. It is possible that increasing the epoch number could solve this issue. The poorer performance in the MLP case (where the performance is spiky in the data), the activation function was sigmoid. It is possible that the activation function could not approximate as well between the combination of stochastic gradient and logistic activation function. The other activation functions did not appear to exhibit the same issue, unless the stochastic gradient descent was paired with a constant learning rate. In the Deep MLP, the hidden layer size did not appear to have a substantial effect on the algorithm performance. However, it should be noted that for the deep MLP case the saved algorithms (best performance in both the training and test cases) did not exceed three hidden layers.

Table 6.4 contains the average algorithm response for determining the different fatigue stages. Unlike the contamination case, the algorithm performance follows a similar expectation to what was found in the bearing damage diagnosis observation. In this instance, the 6206 case did not appear to drop relative to the in the fatigue approximation. One of the possible reasons for the better classifier results could come from the more homogeneous data clusters as a result of the testing. It was noted in the contamination testing that the contamination particles caused the variation in the vibration data. Hence, further overlap could occur between the different speed and damage cases, leading to overlapping data clusters between damage cases.

Figure 6.7 considers the test and train scores for the fatigue case concerning the different parameters used in the analysis. The higher average failure scores showed that the algorithm performance across the different parameters did not vary significantly. The effect of some of the variations seen in the MLP algorithms was reduced. That is more likely related to the changes in the dataset used to perform the training. In the RF and DT cases, the algorithm performance appeared to increase in certain cases, similar to how it occurred in the contamination case. There was still a level of overfitting associated with the algorithm due to the large difference in the test and train scores. However, the effect was reduced in comparison to the fatigue case. From the fatigue case it

Table 6.4: Bearing classifier performance by Average F1 score for determining fatigue stages (Train/Test) during the grid search

Algorithm	6205	6206	6207
Naive Bayes	0.978/ 0.978	0.976/ 0.999	0.9937/ 0.994
Support Vector Machine	0.978/ 0.978	0.976/ 0.999	0.993/ 0.994
k-Nearest Neighbors	0.999/ 0.999	0.999/ 0.999	0.999/ 0.999
Decision Trees	0.997/ 0.999	0.999/ 0.999	0.999/ 0.957
Random Forest	0.997/ 0.999	0.997/ 0.999	0.999/ 0.966
Multilayer Perceptron	0.935/ 0.939	0.936/ 0.938	0.929/ 0.891
Deep Multilayer Perceptron	0.893/ 0.893	0.914/ 0.914	0.919/ 0.918

is possible to consider that the variation is reduced, due to the lack of contamination noise present in the bearing data. This observation was evident based on the identification of single cluster in the spectral augmentation data from Chapter 5.

For the different algorithm cases, the overall importance on the hyperparmeters was reduced as compared to the contamination case, meaning that algorithms performed well overall despite the varying the algorithm parameters. For example in the KNN algorithm, the hyperparemter tuning did not appear to affect the overall performance. In the DT classifier, there was some variation that stemmed from the node splitter function in determining when to continue a branch. However, the performance did not dip below the 95% mark. In the MLP and Deep MLP performance, the stochastic gradient descent algorithm did not perform as well when paired with the inverse scaling and constant learning rates. The adaptive learning rate and adaptive stochastic gradient descent algorithm (adam) provided the best performance. It appears that the neural networks perform best when the non-linear parameters are used, which makes sense cause that is the basis overall design for the algorithms. Additionally, there is a high dependence on the epoch number when the constant rate is used for the algorithms.

6.3.2.1 Individual Speed Case: 1687 RPM

As was mentioned earlier, the different classifiers may perform better when the bearing data are separated based on the different speed classes. The classifiers are tested concerning only the data under the high-speed case (1687 RPM). The bearing labels are assigned in a similar as earlier to test the different classifier approaches. The classifiers were trained under the same grid search as the original classifiers. Table 6.5 contains the classifier performance for the bearing damage
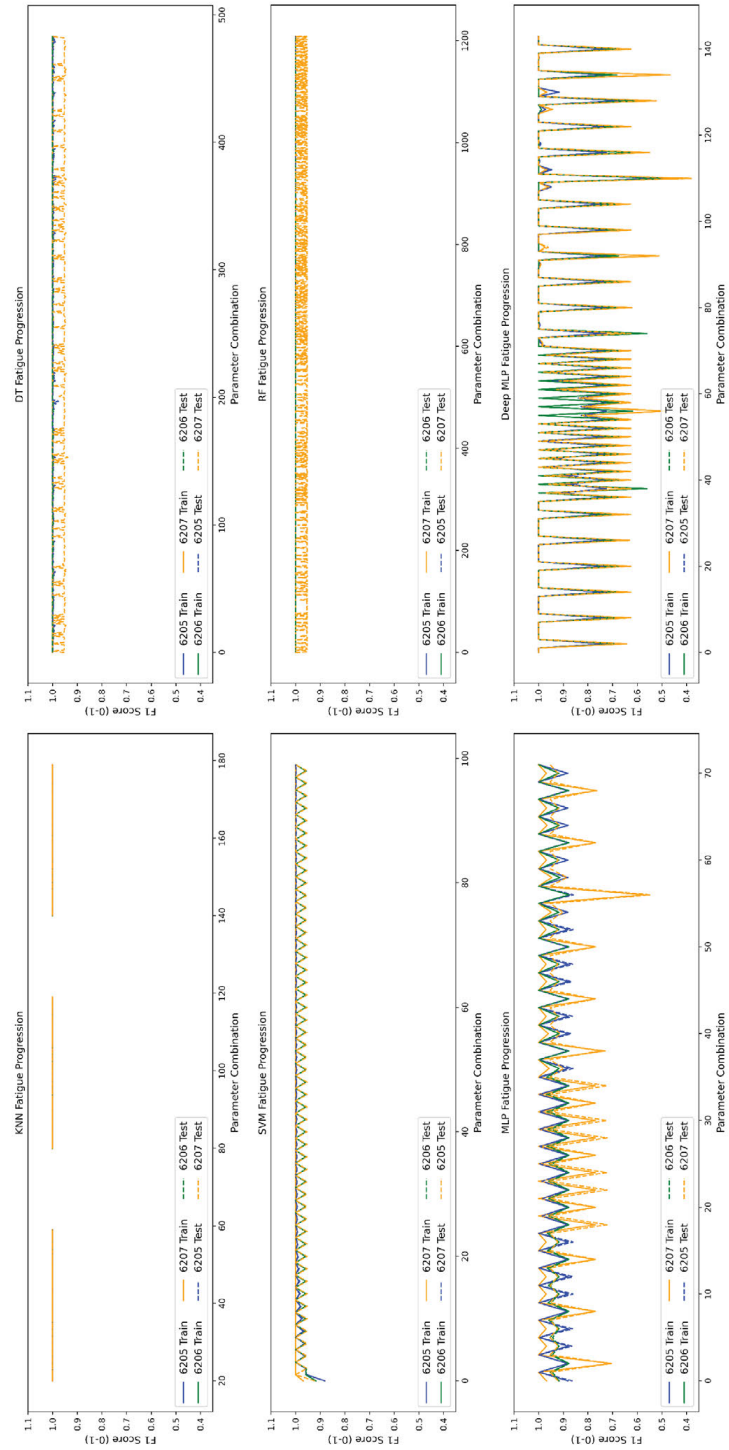
Figure 6.7: F1 training and test scores for each individually tested set of algorithm parameters used in the fatigue cases

328

type for only the data from the 1687 RPM case. In the NB classifier, the damage type increases greatly compared to the all-speed case with the original data. This observation points to potential overlap between the different bearing speed cases. The F1 scores for the other cases remain the same concerning the SVM, KNN, and tree-based models. The MLP and Deep MLP also increase based on this performance as opposed to the single 1687 RPM case.

Table 6.5: Bearing classifiers performance by average F1 score for diagnosing failure type (Train/Test) during the grid search for the high-speed data

Algorithm	6205	6206	6207
Naive Bayes	0.999 / 0.999	0.991 / 0.999	0.999 / 0.999
Support Vector Machine	0.999 / 0.999	0.998 / 0.998	0.999 / 0.999
k-Nearest Neighbors	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Decision Trees	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Random Forest	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Multilayer Perceptron	0.947 / 0.947	0.939 / 0.939	0.965 / 0.965
Deep Multilayer Perceptron	0.876 / 0.876	0.866 / 0.866	0.877 / 0.877

Table 6.6 contains the contamination F1 scores for identifying different contamination stages. There was an increased amount of separability in the Naive Bayes case. Hence, it is possible that by isolating the features relative to the speed, it removes some of the feature dependence between classes. The MLP classifiers did not appear to increase relative to the single speed features. It is possible that these models have parameters that do not approximate well with the data that further skew the results. Hence, it does not appear as though the multiple speed cases would have affected the model training.

Table 6.6: Bearing classifier performance by average F1 score for determining contamination stages (Train/Test) during the grid search for the high-speed data

Algorithm	6205	6206	6207
Naive Bayes	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Support Vector Machine	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
k-Nearest Neighbors	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Decision Trees	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Random Forest	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Multilayer Perceptron	0.888 / 0.888	0.878 / 0.878	0.888 / 0.888
Deep Multilayer Perceptron	0.885 / 0.855	0.876 / 0.876	0.847 / 0.847

Table 6.7 shows the fatigue cases when the classifiers are trained on only the 1687 RPM

data for the different fatigue cases. In this instance, the classifiers performed similar to the test case in the original data case. There was a higher performance in the isolation based on speed which suggests that any misclassifications from the original data case was caused by an overlap in the bearing speed as opposed to the different failure classes.

Table 6.7: Bearing classifier performance by average F1 score for determining fatigue stages (Train/Test) during the grid search for the high-speed data

Algorithm	6205	6206	6207
Naive Bayes	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Support Vector Machine	0.978 / 0.978	0.976 / 0.999	0.993 / 0.994
k-Nearest Neighbors	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Decision Trees	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Random Forest	0.999 / 0.999	0.999 / 0.999	0.999 / 0.999
Multilayer Perceptron	0.913 / 0.913	0.888 / 0.888	0.895 / 0.895
Deep Multilayer Perceptron	0.914 / 0.914	0.893 / 0.893	0.884 / 0.885

In considering only the high speed case for the different bearing failure modes, the classifier performance increase across the board for the 6205, 6206, and 6207 cases for each of the different scenarios. It is possible for the increase to be attributed to the lack of dependence on certain feature types as opposed to when there were multiple different speed cases considered. In the MLP cases, the model appears to perform well only with certain parameters, leading to a lower F1 score in the training and test case. Hence, the similar model performance appears to work regardless of the normalization with respect to speed.

6.3.2.2 Validation Dataset

For the validation cases, a completely separate test set of data not used during the grid search is tested with the best-performing models from the grid search. This test is conducted to see how well the model performs on features not associated with the model training process. The test considers the models that performed the best from the grid search. These models are then tested against a validation set of data that was not seen during the training process. Table 6.8 contains the F1 metric for the different damage types for the best classifiers. As is evident from Figure 6.6, there are multiple instances where the model has a high performance, defined as greater than 95% except in the Naive Bayes case. No modifications or hyperparemeter tuning were conducted, so only the one model iteration was available.

Table 6.8: Bearing classifiers performance by best F1 score for diagnosing failure type

Algorithm	6205	6206	6207
Naive Bayes	0.833	0.962	0.891
Support Vector Machine	0.999	0.999	0.999
k-Nearest Neighbors	0.999	0.999	0.999
Decision Trees	0.999	0.999	0.999
Random Forest	0.999	0.999	0.999
Multilayer Perceptron	0.999	0.999	0.999
Deep Multilayer Perceptron	0.999	0.999	0.999

Table 6.9 contains the model performance for the different contamination stages. The algorithm performance appears to match well with what occurred in the grid search for the classifiers. Depending on the algorithm, the hyperparameters varied with the different bearing sizes. Consistent trends were found in the MLP algorithms, where the identify function was used consistently in each case. Based on the grid search, it should be noted that only the sigmoid activation function performed poorly in combination with other algorithm parameters. Another trend observed was the model convergence with the number of estimators considered in the RF classifiers. It appeared that nothing greater than ten was needed to provide accurate results for the different model structures. The 6205 and 6207 models appeared to perform by the grid search. With the retrained models based on the correct 6206 data, the models performed accurately and as expected compared to the different algorithms.

Table 6.9: Bearing classifier performance by best F1 score for determining contamination stages

Algorithm	6205	6206	6207
Naive Bayes	0.896	0.95	0.925
Support Vector Machine	0.999	0.999	0.999
k-Nearest Neighbors	0.999	0.999	0.999
Decision Trees	0.999	0.97	0.999
Random Forest	0.999	0.999	0.999
Multilayer Perceptron	0.999	0.998	0.999
Deep Multilayer Perceptron	0.999	0.999	0.999

Table 6.10 contains the F1 scores for the best estimators for each bearing size. For each bearing size, the bearing performance performs close to how the model performed in the original grid search. A similar set of hyperparemeters were used as compared to the contamination progression. None of the particular hyperparemeters reflected in the saved models appeared to contradict the

expectations from the grid search.

Table 6.10: Bearing classifier performance by best F1 score for determining fatigue stages

Algorithm	6205	6206	6207
Naive Bayes	0.964	0.914	0.949
Support Vector Machine	0.999	0.999	0.999
k-Nearest Neighbors	0.999	0.999	0.999
Decision Trees	0.999	0.999	0.999
Random Forest	0.999	0.999	0.999
Multilayer Perceptron	0.999	0.999	0.999
Deep Multilayer Perceptron	0.999	0.999	0.999

The validation cases showed good convergence between the different classification tasks. In truth, based on the model performance in the grid search, the classifier performance was expected to remain at a similar level in the validation cases. However, a check was still performed to determine if there were any noticeable deviations, particularly from the contamination data. However, the different models were able to approximate well the variation possibility from additional noise in these particular datasets.

6.3.3   Bearing Health Monitoring: Regression

For the bearing regression, a similar training approach was considered in that a grid search was conducted to assess failure modes and as damage increased. The regression models carried the similar underlying frameworks as those used in the classifier examples. The primary difference between the classifiers and the regression models is the use of predicted value determined as the total vibration energy content of the FFT signal. In this instance, the total vibration energy content was calculated based on the summation of the different portions of the FFT across the different IMF ranges. Hence, the regressor models attempted to fit rules to determine the overall vibration content from the inputs. It should be noted that the error found in the 6206 case for the contamination case for the classifier case did not impact results in the regression case after they were rerun.

Table 6.11 contains the average mean squared error when using all of the training data in one grid search. In this instance, the smaller the scores, the better performing the model. The average mean-squared values appeared to vary much higher as compared to the classifier cases. The primary cause for the high error is the large values predicted between the different cases and the variations

332

seen between the different parameters. Another interesting factor is the amount of overfitting seen in some algorithm cases based on some parameters. For example, in the decision trees and random forest algorithms, the value difference between the training and test cases indicates overfitting for some of the different parameters used in the algorithm cases. The breakdown of where the overfitting occurred is viewed in the grid search for Figure 6.8 and Figure 6.9.

Table 6.11: Bearing regressors performance by average mean squared error for determining damage type stages (Train/Test) during the grid search

Algorithm	6205	6206	6207
Bayes Regression	300 / 500	75 / 110	100 / 120
Support Vector Machine	448.2/643.7	118.8/196.0	76.1/116.9
k-Nearest Neighbors	2.18/89.8	0.764 / 3.57	0.625/407.8
Decision Trees	20.62 / 436.0	17.34 / 65.54	8.745 / 523
Random Forest	22.60 / 387.1	23.88 / 88.28	15.80 / 380.29
Multilayer Perceptron	15729/17457	3844/3976	1500/1599
Deep Multilayer Perceptron	13182 / 14455	7821 / 8652	3978 / 4531

After considering the data in one group, the data were split into their individual damage types of fatigue and contamination and then retrained to see if there was any better performance. Table 6.12 contains the average mean squared error scores for the different bearing sizes and corresponding algorithms for the different contamination cases. Overall the algorithm performance seems to have reduced the amount of overfitting and overall error for each algorithm, except in the Deep Multilayer Perceptron case. It is possible that the noise of contamination stage prevented an accurate prediction of the total energy amount.

Table 6.12: Bearing regressors performance by average mean squared error for determining contamination stages (Train/Test) during the grid search

Algorithm	6205	6206	6207
Bayes Regression	353 / 550	91 / 85	80 / 112
Support Vector Machine	840 / 1300	29 / 45	174 / 229
k-Nearest Neighbors	2.42 / 13.15	0.862 / 0.85	0.617 / 2.699
Decision Trees	37.06 / 93.23	2.08 / 9.33	8.745 / 523
Random Forest	29.76 / 90.39	5.38 / 10.88	19.72 / 82.21
Multilayer Perceptron	10208 / 10472	5668 / 5886	2306 / 2423
Deep Multilayer Perceptron	16060 / 17415	10724 / 12085	5545 / 6418

Figure 6.8 contains different mean squared results for the bearing regressors under different

parameters. There appears to be high overfitting in the KNN, RF, and DMLP cases. The overfitting appears substantial enough to cause a significant amount of error. In the 6207 case, modifying the cost factor for the SVM provided a better result by increasing the value, thereby increasing the upper bound of the norm and allowing for a better maximization between the different classes. In the KNN case, the weight approximation performed better with a distance metric versus a uniform metric.

The DT regressor appeared to perform better when minimizing the L2 loss versus the L1 loss. Minimizing the L2 loss allowed for a better approximation of the algorithm results due to the ability to penalize errors better than the L1 approximation. Outside of that parameter, the algorithm appeared to perform better when using fewer samples to consider a split. For the RF regressors, the influencing parameter was the number of samples used in the split for the different nodes. For the MLP and DMLP, the algorithms performed better when not using the sigmoid and stochastic gradient solver during training. In addition, the algorithm performed better when using more epochs.

Table 6.13 contains the mean squared error for the fatigue cases across the different bearing sizes and algorithms. In this instance, the bearing regressors perform better than the contamination cases. It is possible that the lack of noise in the contamination case failed to skew results. There are still relatively large values in the MLP and DMLP case compared to the contamination case, but that is more than likely caused by the similar set of parameters that force poor results in Figure 6.8.

Table 6.13: Bearing regressors performance by average mean squared error for determining fatigue stages (Train/Test) during the grid search

Algorithm	6205	6206	6207
Bayes Regression	104 / 162	143 / 152	43 / 30
Support Vector Machine	71.8 / 129.4	123.4 / 175.1	35.7 / 73.0
k-Nearest Neighbors	0.062 / 0.271	0.358 / 0.597	0.164 / 0.672
Decision Trees	31.45 / 65.6	19.3 / 34.3	14.27 / 20.61
Random Forest	9.92 / 23.6	56.92 / 56.73	1.44 / 5.09
Multilayer Perceptron	102 / 120	764 / 884	381 / 411
Deep Multilayer Perceptron	488 / 583	1553 / 1785	1247/1495

Figure 6.8 contains the grid search results for the different test parameters for the regression contamination case. Compared to the classifier case, there is a greater variation in the difference in results. In this instance, the negative scores are presented as that was the output of the grid
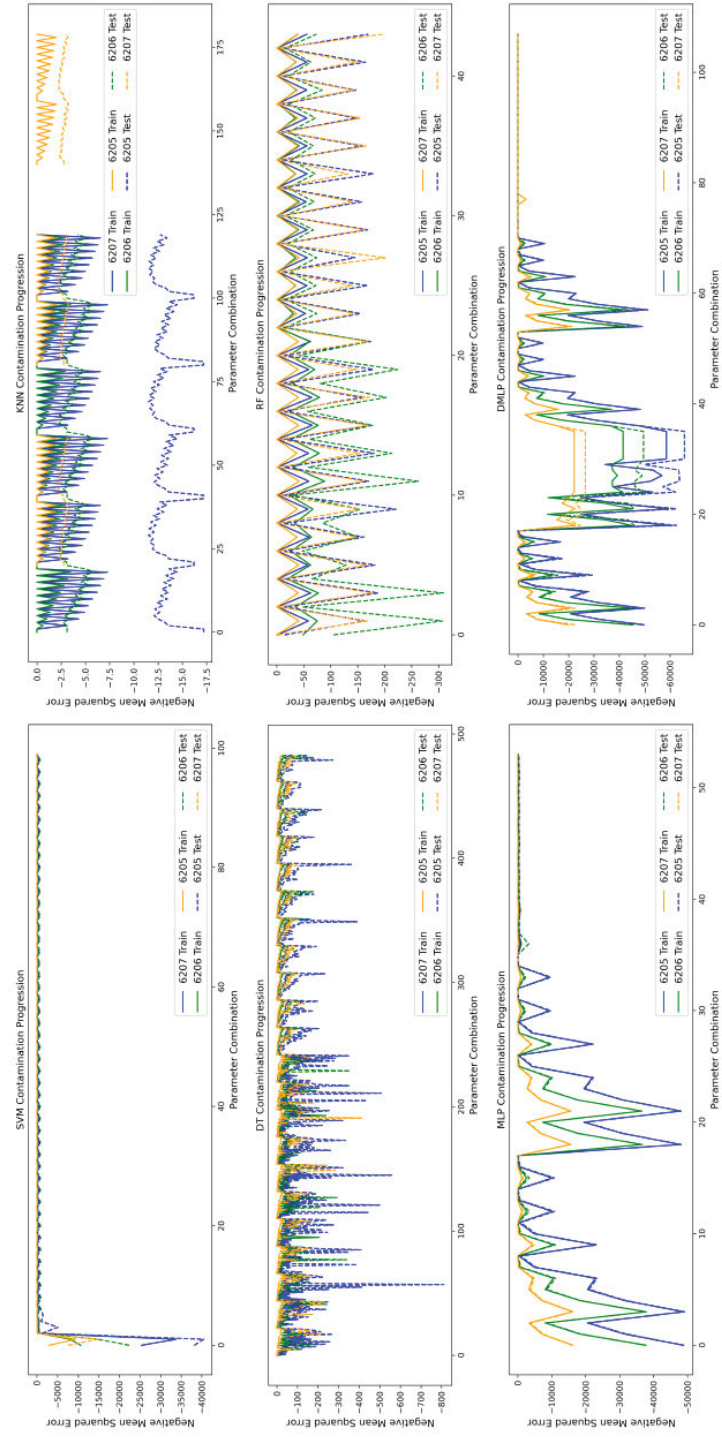
Figure 6.8: Results from the regressor grid search for the contamination case

search loss. It is the same value; just one is negative and the other positive. Similar results are apparent from the regression case and the classifier case. For example, in the support vector case, the algorithm performed very low when considering the lowest cost parameter. That is why there is a huge spike in that initial case. The variations come from the different kernels used, where the RBF kernel performed best. The KNN algorithm had a very low error despite the large change between the different parameters. Using the distance metric to determine weights provides better results than assuming the weights are uniform.

For the DT and RF contamination progression case, the bearing regressors performed better when using the squared error versus absolute error to measure the quality of splits in the tree. The squared error minimized splits related to L2 loss, whereas the absolute error minimized splits related to L1 loss, as mentioned earlier in the damage type considerations. The L2 allows for a faster model convergence with larger penalty factors. The MLP and DMLP algorithms performed poorly using the sigmoid function and stochastic gradient descent solver. The results may point to an instance where the vanishing gradient descent lowers the ability of the algorithm to update.

In essence, the regressor algorithms exhibit greater variance in results during the grid search than the classifiers. The primary reason is the variation in the results due to the contamination data. The additional noise caused by those elements appears to stem from the variation in the particle contact. Despite the variation, some parameters provided very low error results, leading to the possibility of low overfitting and good prediction of the regression results relative to the different health stages during the validation testing.

6.3.3.1   Regressor Validation

Like the classifier cases, the regression models were tested with a validation dataset not used in the grid search. Table 6.14 contains the validation results for the 6205 case for the respective damage type, contamination progression, and fatigue progression. Compared to the test cases, the validation case appears to have an overall low error, except in the Naive Bayes contamination case. The reason for the high error could stem from the variation seen as a result of the contamination case and the lack of ability to approximate unseen data.

However, the other algorithms perform relatively well. Figure 6.10 contains the validation case for the different damage types concerning the different regression models in the damage type data case. In this instance, despite the relatively low error, the models approximate the predicted

336
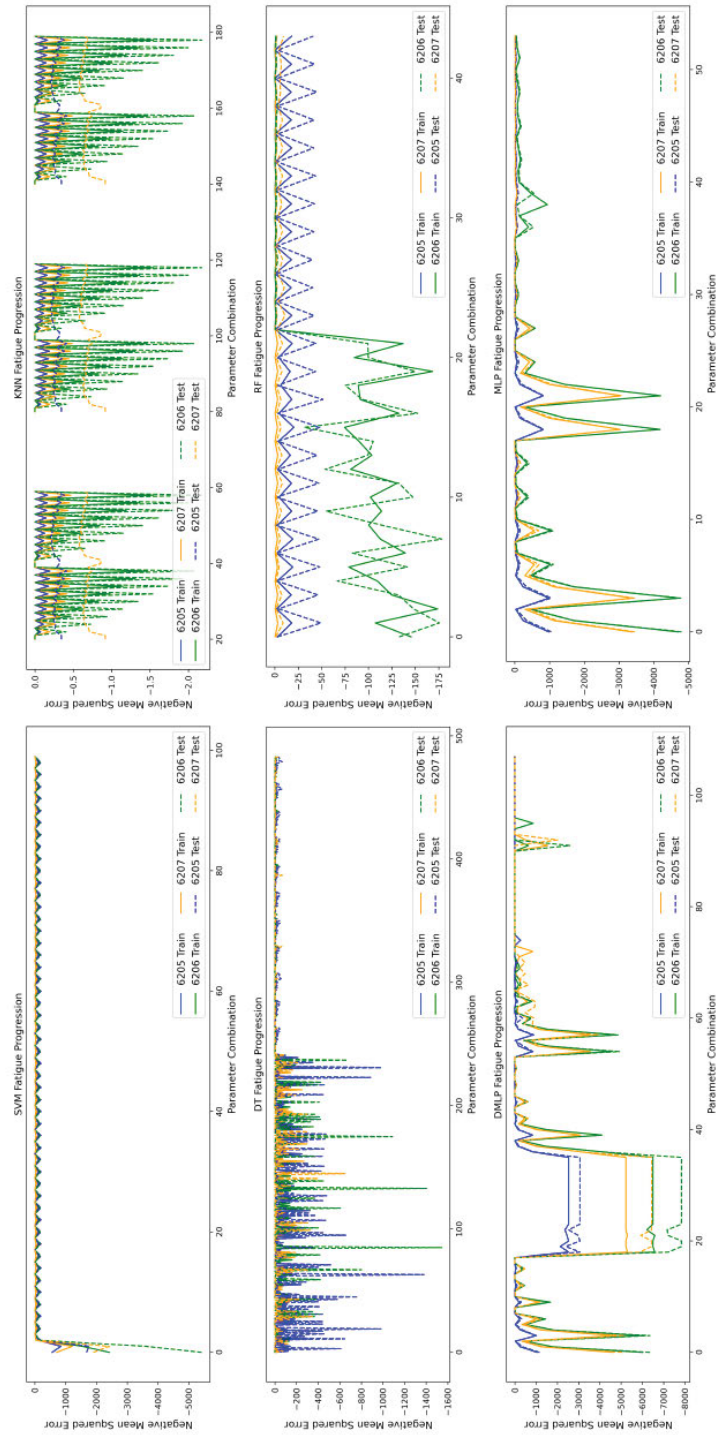
Figure 6.9: Results from the regressor grid search for the fatigue case

Table 6.14: Bearing regressors performance by average mean squared error for determining fatigue stages (Train/Test) during the grid search

Algorithm	Damage Type	Contamination Pro.	Fatigue Pro.
Bayes Regression	20.54	659.00	111
Support Vector Machine	20.53	48.60	1.24
k-Nearest Neighbors	15.11	16.48	0.35
Decision Trees	104	27.37	1.16
Random Forest	26.21	23.36	21.91
Multilayer Perceptron	20.47	35.11	0.94
Deep Multilayer Perceptron	20.47	24.94	0.97

cases relatively well for the 6205 case, the data along the y-axis that increases exponentially is the contamination vibration energy Code 0 case. That appears to have the most trouble in the approximation relative to the different IMF cases. The relative error remains low for the different validation cases. Overall, the models approximate the unseen data from training relatively well by predicting values in the center of certain clusters. However, the difference in values could lead to possible errors if further outliers are recorded due to further damage.

6.3.4 Bearing Health Monitoring: LSTM, HMM

The same input features were used for the LSTM models as were considered for the previous machine learning models. The labels were assigned based on the total vibration content within eah cell. Three different models were trained for each bearing case: (the damage type, the contamination prediction, and fatigue prediction. The LSTM models were trained on a two-part grid search. The first part of the grid search sought to optimize the number of LSTM cells per layer and the number of layers. A second search was then conducted to determine the effect of epochs and batch size on final model performance. The algorithm did not see a significant change after adjusting the number of epochs and the batch size. Hence, only the results from the first grid search were considered. Table 6.15 contains the results from each model based on the mean squared error.

Table 6.15: Bearing LSTM performance by best mean squared error score for diagnosing failure type (Train/Test)

Case	6205	6206	6207
Damage Type	0.823 / 20.39	4.197 / 12.07	2.317 / 19.04
Contamination	16.04 / 29.06	18.73 / 79.75	0.402 / 17.65
Fatigue	0.311 / 0.513	0.919 / 164.3	0.033 / 1.312
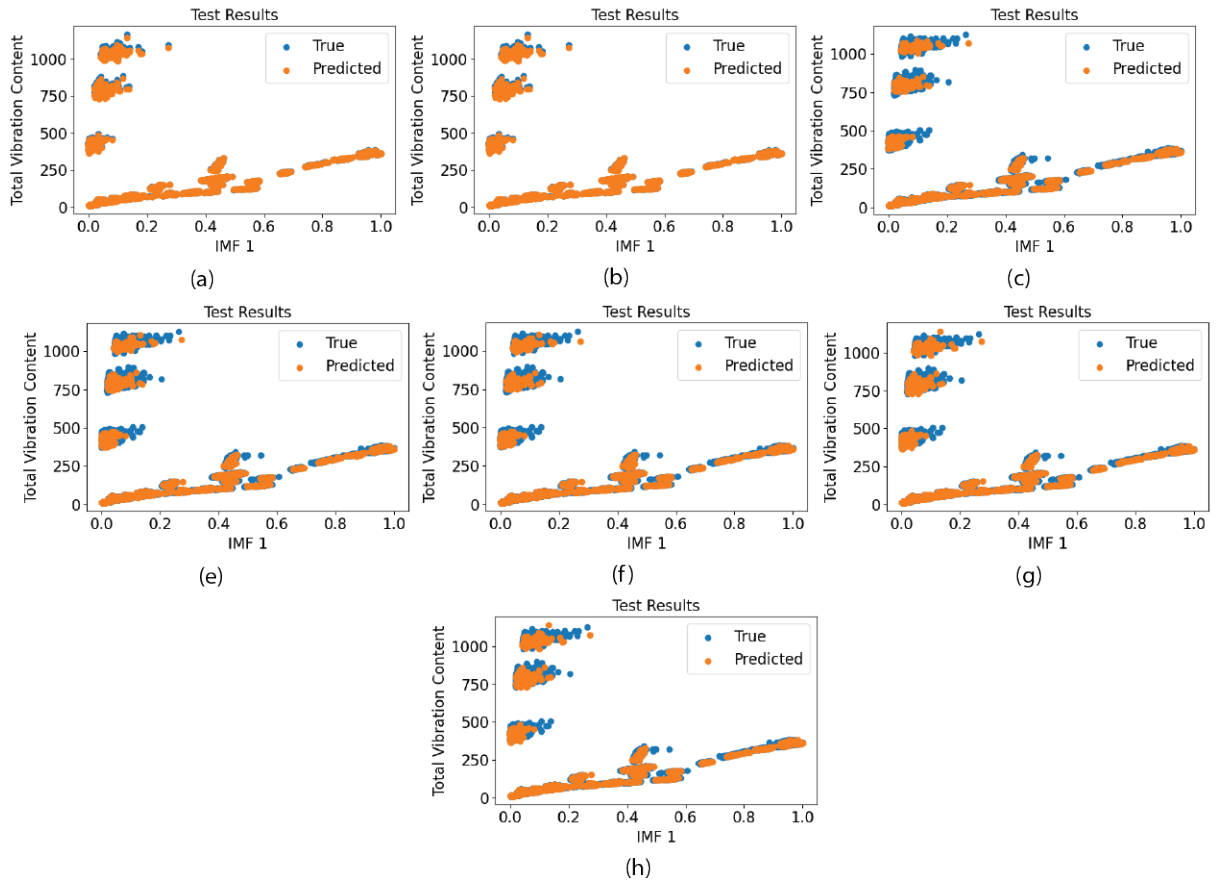
Figure 6.10: Model comparison between true and predicted results for the (a) Naive Bayes, (b) Support Vector, (c) kNN, (d) Decision Tree, (e) Random Forest, (f) Multi Layer Perceptron, and (g) Deep Multi Layer Perceptron

The values above appear large in terms of error compared to the classifier scores. However, there is a higher degree of expected overlap in comparing the true and predicted values for the regression. Figure 6.11 shows the comparison of true versus predicted values in the case of fatigue and contamination combined (Figure 6.11 (a) - (c)). Despite the large error in these instances, the model can adjust depending on the speed and damage case. The predicted points show that the difference primarily comes from the difference scale between the true and predicted cases. Even when the model has difficulty addressing the change in points, such as in Figure 6.11 (d), there is still a closer convergence that allows for the prediction of most cases. It is odd to see, though, that the 6206 Fatigue case (Figure 6.11 (d)) did not perform as well as the other bearing cases. It is possible that further model tuning could correct the overall performance to accommodate that final failure case.

339
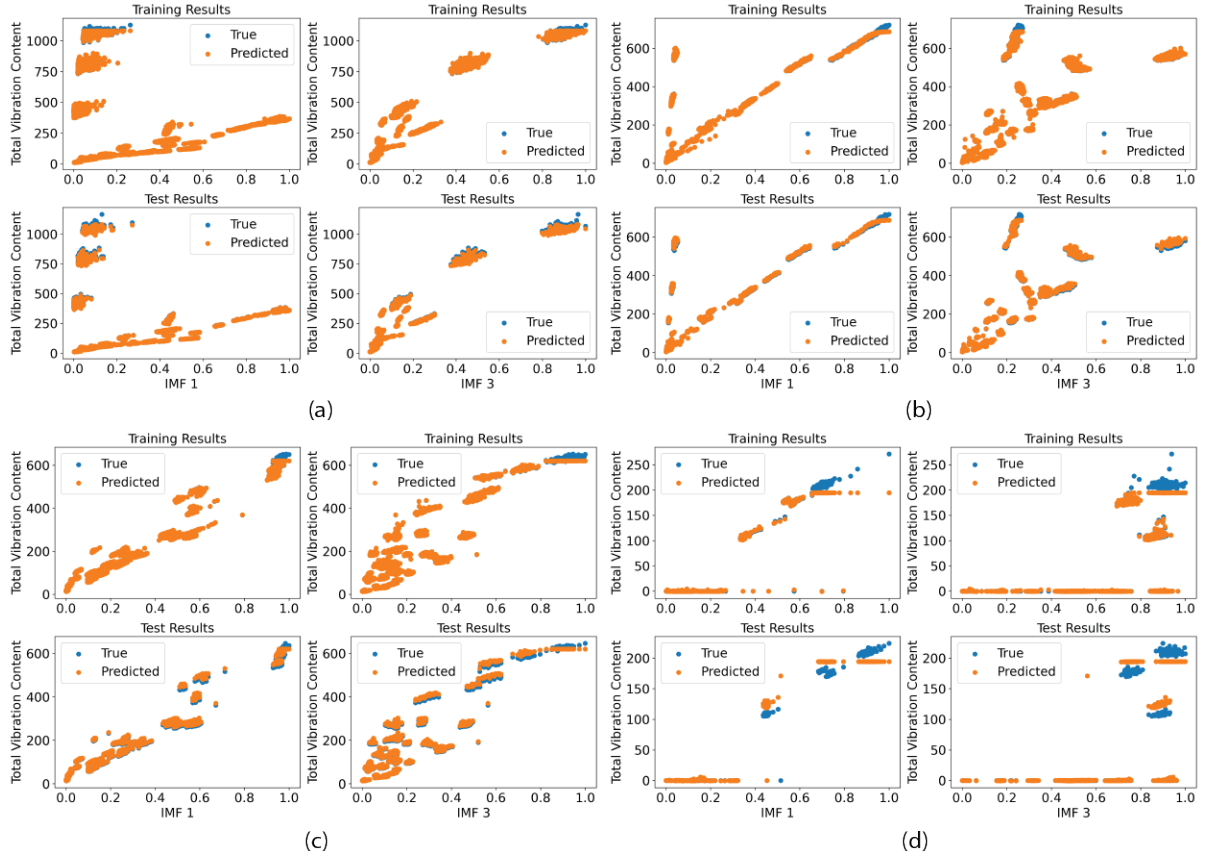
Figure 6.11: LSTM model results for all contamination and fatigue data for the (a) 6205 case, (b) 6206, (c) 6207 case, and the lowest performing model (d) the 6206 fatigue case. Each figure has the training and test result compared to the values from IMF 1 and IMF 3

Overall, though, the LSTM model can predict the total energy content of the bearing regardless of the damage type and speed cases. Interestingly, two general trends are shown in the 6205 case and 6206 case for IMF 1. Based on viewing the samples, the associated instance of where the data remains close to 0 on the x-axis but increases along the y-axis points to contamination damage. The more general line is attributable to damage associated with fatigue cases. In the 6207 case, that change in damage is not as readily seen in the IMF 1 case. However, the general trend of exponential increase in the IMF 3 case is seen as damage is applied and corresponds to increases seen in the 6206 and 6205 cases. It should be noted that the x-axis vibration is normalized based on a Min-Max scaler. Despite that normalization, the underlying information remains the same and there appears to be an exponential curve associated with the increase in bearing vibration along the IMF 3 case as the bearing vibration energy increases.

340

In contrast to the prior classifiers, regressors, and LSTM, an unsupervised classifier approach was conducted for the HMM to determine if the unsupervised classifier approach could determine any particular trends without the need for labels. The unsupervised learning approach was also conducted to determine the proper probabilities. These are known, given the fact that the data was collected in a lab. However, in a production environment, these parameters may not be known, hence the adoption of the unsupervised learning approach. The feature configuration was the same between the supervised models and HMM.

Table 6.16 contains the HMM parameters and log probability scores for the different hmm classifiers. For the HMM classifiers, the number of hidden states was constrained based on the health stages and the corresponding speed cases. There were eight different health stages (four contamination stages, three fatigue stages, and one baseline stage) and three different speed cases. Hence, a maximum number of 30 hidden components were considered. The log probability score for each model was the quantity maximized to determine the goodness of fit. Given the number of samples provided in both the training and testing case, the log probability scores approximately represent the ratio of points used for the training case (80%) and (20%) for the testing case. The log probabilities represent this split in every case except the 6206 data, where the values represented a case of 85% training score and 15% of the test score. Hence, the test data may have had a poorer model performance.

Table 6.16: Bearing HMM Characteristics and Log Probabilities for each set of feature data

Case	6205	6206	6207
# of hidden states	16	15	15
Train Probability	349404	265243	227723
Test Probability	86442	47871	53907

From the number of hidden states, it is possible that some of the different health stages and speed cases were lumped together. Figure 6.12 shows the approximation of the hidden states for the different bearing cases compared against the original data labels in the bottom three images for each column. In each case, the colors appear to represent different classes. In each instance, teh amount of separation appears to determine what is a new class and what fits with an existing label.

In each instance, the baseline data appears represented by three colors, signifying the different speed cases. There is at least one severe contamination case found in each instance, where the different clusters represent the breakdown of speed. For the 6205 and 6206 case, the clusters

341
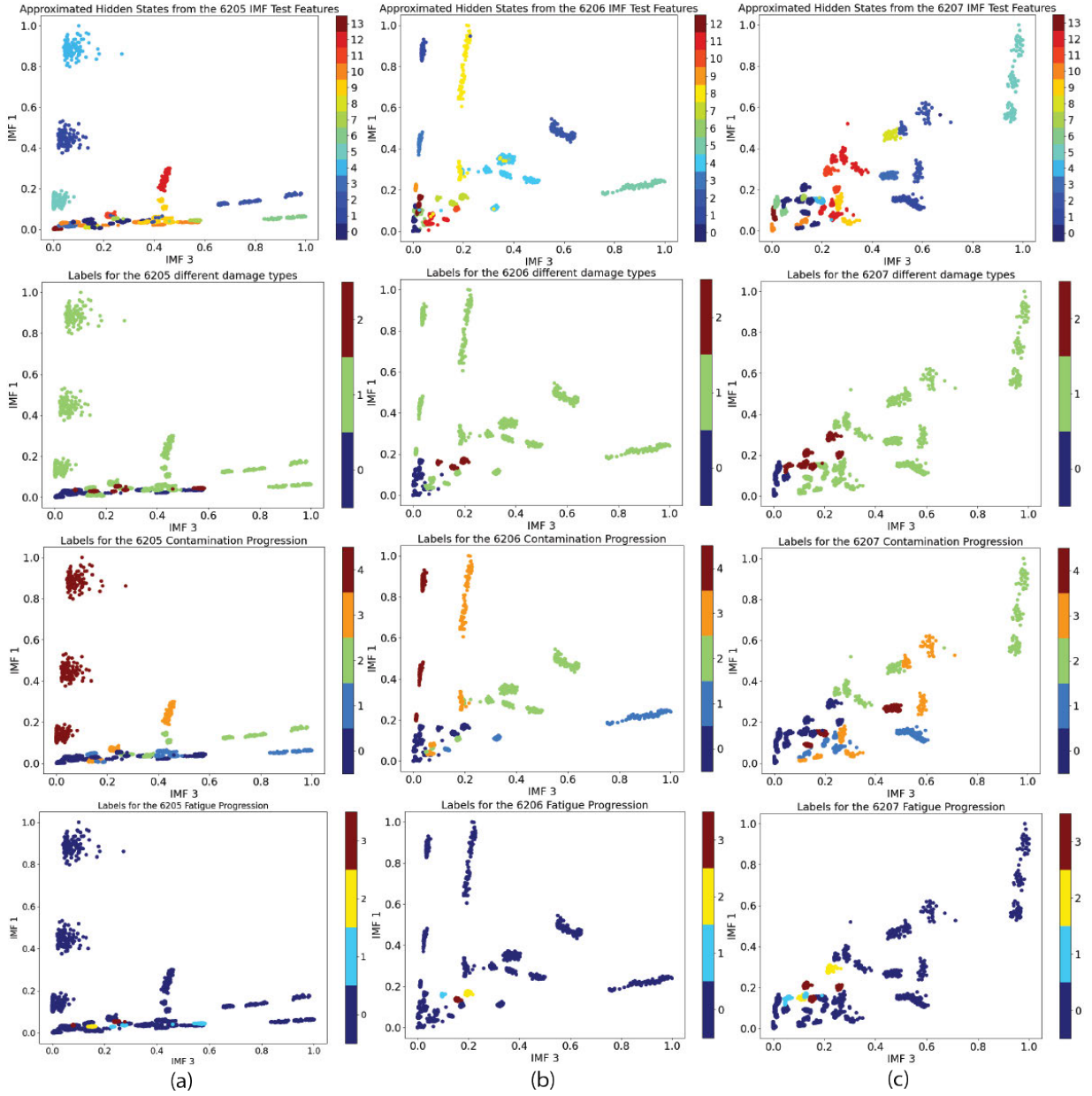
Figure 6.12: Columns for each bearing size: (a) 6205, (b) 6206, and (c) 6207. Each column has the HMM results, followed by the labels for the different damage types, the contamination progression, and the fatigue progression

appearing closest to the y-axis appear to denote a variation in speed for one of the contamination cases. For the fatigue case, the data appears to be too intermixed with the contamination and baseline case for the determination of separable data classes for the 6205 and 6207 cases. For the 6206 case, interestingly, the fatigue data appear to all be labeled into one group. The close approximation could stem from similarities across other IMF sets.

6.3.5 Key Takeaways: Original Data Training

The following analysis provided a baseline for testing the different algorithms. From this point forward in the research, the models generated are compared using the augmented bearing data to train and diagnose potential failures within the system. As the original data shows, the overall data amount provides a high accuracy of different failure modes and health stages. It is the goal that the algorithm metrics remain similar a different data types are used to diagnose and predict the different bearing defects.

6.4 Model Performance: Transfer Models

Before applying the augmented bearing data to different models, another algorithm test was conducted to see if the model parameters from a different bearing size produced similar results for the bearing results. In this instance, the bearing classifiers are tested by taking the 6205 data and applying the models from the 6206 and 6207 cases. In this instance, the labels are considered the same between the different bearing damage classes. Table 6.17 considers the score from the different damage type classifiers. Overall, the performance of these classifiers outside of the Naive Bayes case follows the validation case data. The separation in the data may appear unrealistic at first glance; however, one should consider that in most applications, defects consider other noticeable occurrences, such as sound and changes in temperatures. While these changes may appear to change over time, they are prominent indicators for changing equipment conditions and operation. Similarly, while there is a large degree of separation in the vibration features, there is still a clear indication of changing conditions as a result.

Table 6.18 denotes the algorithm performance for the contamination classifiers from the 6206 and 6207 cases trained using the 6205 bearing data. Both models appeared to perform well, except again in the Naive Bayes case. The most probable cause is the assumption of independence

Table 6.17: F1 score for the damage type classifiers utilizing the 6206 and 6207 models with the 6205 features

Model	6206	6207
Naieve Bayes (NB)	0.833	0.833
Support Vector Machine (SVM)	1.0	1.0
kNN	1.0	1.0
Decision Trees (DT)	0.998	1.0
Random Forest (RF)	1.0	1.0
Multilayer Perceptron (MLP)	1.0	1.0
Deep Multilayer Perceptron (DMLP)	1.0	0.993

between features. In truth, the similar classifier performance supports the particular trends observed in the parameter grid searches from Figure 6.6. Overall, the performance remains strong despite the different parameters. The scores in Table 6.19 support a similar trend for the fatigue case.

Table 6.18: F1 score for the contamination progression classifiers utilizing the 6206 and 6207 models with the 6205 features

Model	6206	6207
Naieve Bayes (NB)	0.889	0.898
Support Vector Machine (SVM)	0.998	1.0
kNN	1.0	1.0
Decision Trees (DT)	1.0	1.0
Random Forest (RF)	1.0	1.0
Multilayer Perceptron (MLP)	1.0	1.0
Deep Multilayer Perceptron (DMLP)	1.0	1.0

Table 6.19: F1 score for the fatigue progression classifiers utilizing the 6206 and 6207 models with the 6205 features

Model	6206	6207
Naieve Bayes (NB)	0.96	0.96
Support Vector Machine (SVM)	1.0	1.0
kNN	1.0	1.0
Decision Trees (DT)	0.991	1.0
Random Forest (RF)	1.0	1.0
Multilayer Perceptron (MLP)	1.0	1.0
Deep Multilayer Perceptron (DMLP)	1.0	1.0

As an understanding of where the potential overlap occurred, Figure 6.13 provides the confusion matrices for the 6206 and 6207 Naive Bayes classifiers concerning the different cases the classifiers were tested for. The classifiers appeared to perform similarly in each case, which makes sense given that the model did not change between the different data cases. In the damage type plots (Figure 6.13 (a) for the 6206 case and Figure 6.13 (d) for the 6207 case), the bearing data

appeared to approximate the contamination cases in either baseline case or the fatigue case. It is possible that in the lower contamination case (Code 20) and cleaned contamination case (Code 0), there was enough samples that were intermixed with the data during certain speed classes. That appears in the case of the HMM plots from Section 6.3.4.
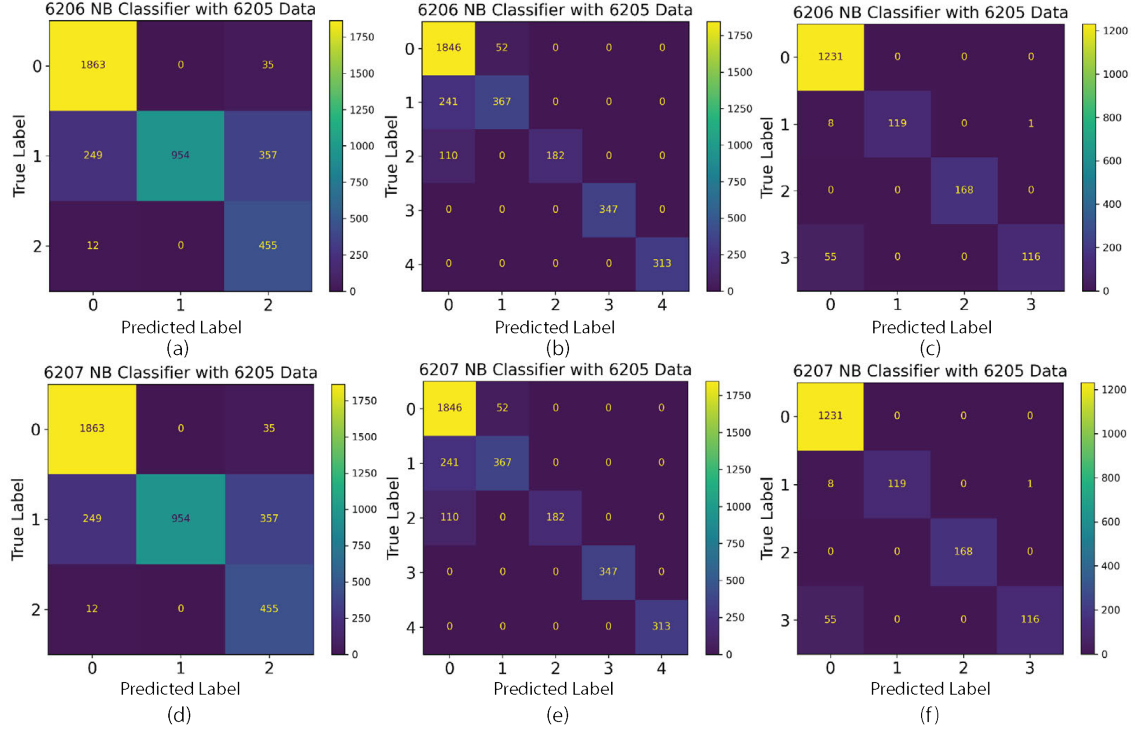
Figure 6.13: NB confusion matrix from the different classifier types applied to the 6205 features: (a) 6206 damage type classifier, (b) 6206 contamination progression classifier, (c) 6206 fatigue progression classifier, (d) 6207 damage type classifier, (e) 6207 contamination progression classifier, and (f) 6207 fatigue progression classifier.

The error appears to occur for the contamination case based on the lower contamination cases. These contamination points are likely the result of misclassification with some of the higher-speed cases. It is possible that the Naive Bayes case found some similarity across the lower frequency cases that outweighed the probability of these cases being determined by their upper frequencies. It is also possible that due to the wide dispersion of contamination features across the different classes, the algorithms for the lower contamination cases and the high-speed baseline cases were intermixed. In the fatigue case, the single clusters formed due to the lack of contamination reduce the error between the different cases. That possibly explains why the classifiers for the fatigue case did not perform as poorly in comparison to the contamination cases.

Table 6.20 contains the mean squared error for the 6206 and 6207 data trained on the 6205 data. Similar to the classifiers, there is not a discernible change in the prediction cases compared to the error results in Table 6.11. It appears that for the different bearing sizes similar parameters will work one each set of data. The largest difference in the DT algorithms was the use of the splitter algorithm to evaluate the quality of the split. In this instance, it was interesting to note that the mean absolute error performed better as compared to the mean squared error. That was unexpected as it is generally the mean squared error (L2) that is selected to approximate the regression output. It could be that once the maximum depth is of a significant size that the criterion for assessing the split does not influence results heavily. Despite that change in value though, the lower error values do correspond to what is expected from the original data approximations.

Table 6.20: Mean squared error score for the all regressor cases utilizing the 6206 and 6207 models with the 6205 features

Model	6206	6207
Bayes Regression	20.548	20.548
Support Vector Machine (SVM)	20.525	20.525
kNN	15.981	14.984
Decision Trees (DT)	40.037	28.56
Random Forest (RF)	19.951	21.863
Multilayer Perceptron (MLP)	20.550	20.549
Deep Multilayer Perceptron (DMLP)	20.807	20.737

Table 6.21 considers the different LSTM models transferred between the different bearing sizes and assessed based on the validation set. For the 6205 case the models performed poorly when using the features from the different corresponding training cases. The difference could stem from the lack of nodes in each layer as compared to the 6206 and 6207 case. In the 6205 case, 16 nodes were needed to approximate the data, whereas in the 6206 and 6207 case only nine nodes were needed in both bearing size cases. The need for more nodes could stem from the need to flexibly transform the data in a manner that allows for the best approximation of vibration content. In the case of the 6206 and 6207 data, the amount of nodes needed was not as high. In the 6026 and 6207, the error values appear to approximate the algorithms well overall.

Table 6.21: LSTM results using features and other corresponding models

Model	6205 Model	6206 Model	6207 Model
6205 Features	-	3764	5650
6206 Features	48.50	-	131.074
6207 Features	53.14	24.46	-

The unsupervised learning algorithms (HMM) were not considered for the transfer learning approximation. Transfer learning usually incorporates elements extracted from supervised learning versus unsupervised learning. From the different algorithms, the bearing data appears to transfer between the different bearing sizes well except in the 6205 LSTM case. Despite that deviation, the trasnference of model learning is possible between bearings of similar size under the same conditions.

Data from a production system motor were considered to see if the model parameters were useful in a real world system. Two collections are available concerning the production equipment. The first collection was collected one week before the failure occurred. Prior to that failure a historical analysis found that anomalous vibration readings were noticed at least a month in advance. There is no knowledge related to the failure, other than that the bearing failed. After the failure, a second collection was collected to verify that the system was back to a baseline state and was collected three days after. For the purpose of this analysis, the data collected prior to the failure were labeled as failure data, and the data collected after were considered baseline data.

The data were preprocessed and transformed into features in a similar manner as conducted from the test stand data. The models created from the test stand were tested and trained with the data from a production system motor. A similar split was used as was considered in the test system scenario. Table 6.22 contains the classifier performances for the model cases from the motor defect scenarios. The model performed well in determining the differences between healthy and failed bearing data, when labeled as described above. Hence, each model scenario can determine differences equally well when considering a binary case under one speed condition. This observation holds true when compared to the test stand data.

Table 6.22: F1 score from classifier performance from a live motor in industry

Model	6205	6206	6207
Naive Bayes	0.999	0.999	0.999
Support Vector Machine (SVM)	0.999	0.999	0.999
kNN	0.999	0.999	0.999
Decision Trees (DT)	0.999	0.999	0.999
Random Forest (RF)	0.999	0.999	0.999
Multilayer Perceptron (MLP)	0.999	0.999	0.999
Deep Multilayer Perceptron (DMLP)	0.999	0.999	0.999

Table 6.23 shows the results when considering the regression results as oppose to the classification results. In comparison to the previous regression results, these results seem a lot higher in comparison to test stand. The differences between the classifiers and regressors is a common occur-

rence when determining a binary state versus predicting a discrete value. The unexpected increase in error for the regression analysis is surprising to see and could point to not enough samples for an accurate prediction. The ratio of samples from the real world to the original world data is that there were at least 100x more samples from the test stand versus the real world scenario. However, this occurrence is expected and highlights the reason for this research.

Table 6.23: Mean square Regressor performance from a live motor in industry

Model	6205	6206	6207
Bayes Regression	1162.8	1162.8	1162.8
Support Vector Machine (SVM)	724.60	1138.3	1018.4
kNN	916.96	989.45	920.72
Decision Trees (DT)	1076.2	952.75	57656.8
Random Forest (RF)	565.72	715.40	607.13
Multilayer Perceptron (MLP)	687524	712335	730816
Deep Multilayer Perceptron (DMLP)	49974.9	168281.4	36240.6
LSTM	340944	1718926	121409

Figure 6.14 displays the predicted values when using the different regression models to determine the total vibration energy content. The variation in the error values is more apparent when considering the scale on the y-axis. However, it appears that the most probable cause of variation is due to the low amount of samples as was mentioned earlier when assessing the error values. The trend was observed more so in the neural networks, where the trends appeared to try and conform to the model results, but failed to converge on the values themselves. It is possible that additional epochs would be needed. In the 6207 case for the decision tree, the model completely failed to converge to the expected data values. Hence, there are certain parameters that would provide a better result concerning the data. In this instance, the criterion for measuring the quality of a split appears to suggest the difference.

The transfer learning section was meant to highlight whether it is enough to transfer only the model between different scenarios. For the test stand features generated, the models appear to perform well and support the use of models from other datasets to train and predict trends on similar datasets. However, there are still instances where the model transfer appears to perform poorly as noted in the LSTM regression for the 6205 features using the 6206 and 6207 models. As noted earlier, additional nodes may be needed to approximate an unknown nonlinear component not seen in the 6206 or 6207 data. A production example was considered from real world scenario where a failure was identified. The classifiers performed well in determining the difference between
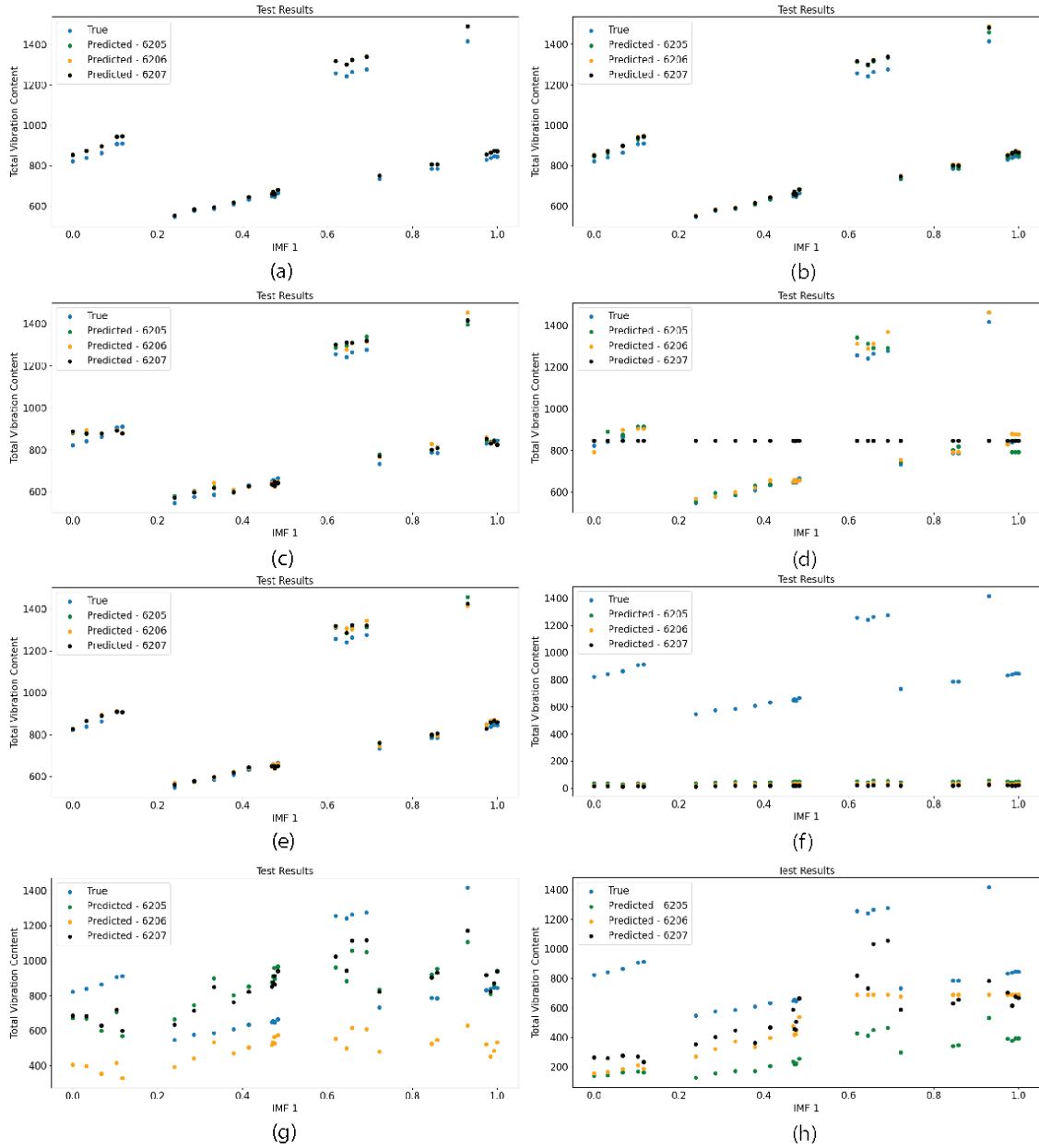
Figure 6.14: Predicted motor failure based on the input parameters and under the different model considerations from the different bearing sizes: (a) Naive Bayes, (b) SVM, (c) KNN, (d) DT, (e) RF, (f) MLP, (g) DMLP, and (h) LSTM.

the baseline and healthy case. However, the regression struggled to perform well overall, especially in the neural network situations. It is possible that the lower amounts of training data may have caused errors in the approximation. There were some model errors highlighted as shown in the 6207 DT model application (Figure 6.14 (d)). Hence, overall, to increase model accuracy, scaling

349

data from a test sytem scenario to the production system scenario may in turn increase the model performance between each system.

6.5 Model Performance: Augmented Data

In the augmented case, the best model from the original training data was used to see how well the augmented features were classified. The two different augmented measures from Chapter 5 (Physical Scaling and Spectral Augmentation) were consdiered as the methods for creating augmented data. For this analysis, augmented data is any data created based on the use of experimental data form either the source application only or from the target and source application. For each augmented method, the models from the original data training case were used to determine if the augmented data was similar to the experimental data. The goal was to evaluate how well the augmented data performed under model characteristics from the original data. If the models were able to accurately classify or predict the data measure, then the augmented data could be considered to hold similar training applications to the original data.

Three different experiments were considered initially. The first experiment considered if the data used to train the different algorithms only came from the experimental data collected in Chapter 4. The model assessment occurred with both the augmented and true data that was left out of the training process. The second experiment considered if the training data set contained equal amounts of both original and augmented data. The assessment occurred with both data types. The third case considered if the models were trained using only the augmented data. The model assessment was then conducted using any leftover bearing augmented data and the true data leftover.

6.5.1 Augmented Dataset: Physical Scaling

The physics scaling process was tested based on the assumptions considered originally in the process description for Section5.1.5. The data created in that section are utilized to verify how they perform in the model application. Hence, only the fatigue case is considered in this application for the 7 mm (Inner Race) I.R. defect and the 1687 speed case. The decision to test this particular defect case versus another was due to the frequency response seen in the model output that appeared to match close to expectations in relation to the experimental data. Table 6.24 contains the breakdown of the data used in the testing for the physical scaling process.

Table 6.24: Breakdown of data forming the augmented and true cases for testing in Section 6.5.1 for the 6205 and 6207 bearing sizes. PS means physics scaling process.

Data Types	True	Augmented
Healthy	Train Baseline (50% 3K, 50% 4K, 50% 5K, 50% 6K 6207), Test Baseline (50% 3K, 50% 4K, 50% 5K, 50% 6K 6207)	-
Failure	Train Fatigue I.R. 7 mm (12K 6205 (only for augmentation)), Train Fatigue I.R. 7 mm (50% 12K 6207), Test Fatigue I.R. 7 mm (50% 12K 6207)	PS Train Fatigue I.R. 7 mm (50% 12K 6207 Train Fatigue), PS Test Fatigue I.R. 7 mm (50% 12K 6207 Train Fatigue)

The bearing data for the I.R. 7 mm defect case comes from the 6205 and 6207 bearing case, which was tested in Section5.1.5. Based on the results from the physical properties, the time series method where the RMS does not scale the bearing data is selected because the method has the most similarity between the different feature groups based on Levene's test. No augmented healthy data is created as a result of the process. For Case 1, only the real data is used in the training process. For Case 2 only, the training data for the failure case is split equally between the true data and the augmented data from the PS case. For Case 3, only the augmented data are considered for the training case (PS I.R. 6207 Augmented Data). The test cases consider all of the remaining available data.

6.5.1.1   Classification Results

Table 6.25 contains the F1 test scores for the different classification. For the Case 1 and Case 2, the performance is near perfect for the different classifiers. It is possible that based on the results the true failure data approximated the augmented failure data better than the baseline data. However, with the Case 3 results, the opposite case does not appear true. In this instance, the results appear to support that there is too much overlap in certain features that confuse the algorithm to select the wrong class. The kNN does perform the best out of the different cases. It is possible that the reason for how well the algorithm performs is due to the separation in some of the features. In the other algorithms case, such as the SVM or Decision Tree, the algorithm trained rules man not be able to train as well due to the rules formed during the training process.

Figure 6.15 shows the different IMF feature groups for the different data cases: Baseline, Augmented Failure, and True Failure for the 6207 case. From the data cases, it is possible to see that there is some overlap in the data, particularly for the IMF 1, 4, 5, and 7 cases. However, there

351

Table 6.25: Classification F1 scores for the different training cases on the 6207 data

Model	Case 1	Case 2	Case 3
Naieve Bayes (NB)	0.999	0.999	0.622
Support Vector Machine (SVM)	0.999	0.999	0.693
kNN	0.999	0.999	0.850
Decision Trees (DT)	0.999	0.999	0.716
Random Forest (RF)	0.999	0.999	0.746
Multilayer Perceptron (MLP)	0.999	0.999	0.844
Deep Multilayer Perceptron (DMLP)	0.999	0.999	0.779

is clear separability for the IMF 3, and 6 cases. Even in the cases for IMF 4, 5, and 7, thee are distinct differences between the data based on the mean point and the variances.
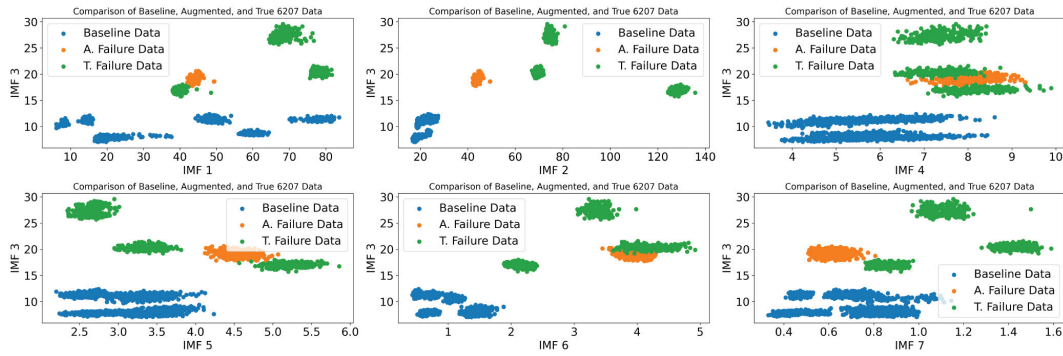
Figure 6.15: Difference in features across the three different types of data: Baseline 6207, Augmented 6207, and True 6207.

Figure 6.16 contains the breakdown of feature importance for the different training cases to try and identify what features are causing the correct classifications versus the incorrect classifications. For Case 1, the most important feature is the IMF 2 feature. For the Case 2, the most important feature is the IMF 3 feature. For the Case 3, the most important feature is the IMF 6 feature. Based on the correlation of this Figure 6.15 and Figure 6.16, it is evident that the variation appears to come from IMF 4 that causes the misclassifications based on the variance associated with that feature. IMF 1 and 5 are not even considered during the decision-making process as a result. Hence, eliminating those four features would potentially increase the algorithm's accuracy.

Despite, the poorer classification results for Case 3, under the consideration of Cases 1 and 2, the classifiers did perform as expected in separating between the baseline conditions and the failure conditions. It is possible to utilize the augmented failure data to enhance possibly an existing dataset. However, it is not yet enough to completely replace the need for true training data. It could
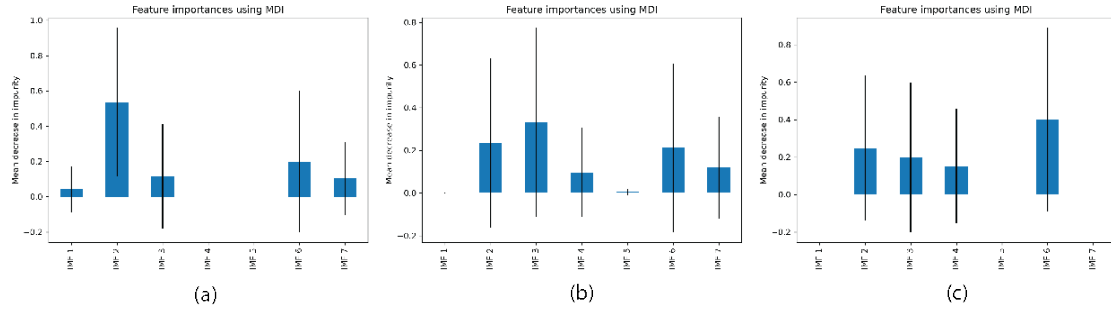
Figure 6.16: Feature importance for the three different cases: (a) Case 1, (b) Case 2, and (c) Case 3 RF classifier feature importance

be possible that the classifier performance would improve by removing certain features as shown in Figure 6.16. However, the removal of features would have to be on a case-by-case basis due to the possibility they contain relevant information regarding the system's physics and health.

6.5.1.2 Regression Results

Table 6.26 contains the regression scores for the different regressors from the original data training. Similarly to the classifiers, the regressors did not perform well in Case 3 for the training data. However, an interesting observation was noted for the SVM, kNN, DT, and RF regressors, where the mean-squared error appeared to decrease as the augmentation data was added. The expectation of seeing the increase in the training score stems from a better ability to approximate variation in the data from similar but different data types. Based on the scores, though, an element of true failure data is still needed in the verification process.

Table 6.26: Regressor mean squared error scores for the different training cases on the 6207 data

Model	Case 1	Case 2	Case 3
Bayes Regression	1.64	2.35	524.04
Support Vector Machine (SVM)	53.09	15.76	1415.96
kNN	162.06	58.511	939.08
Decision Trees (DT)	164.70	61.29	2227.97
Random Forest (RF)	111.78	49.79	1101.77
Multilayer Perceptron (MLP)	2.39	4.9114.911	660.67
Deep Multilayer Perceptron (DMLP)	1.63	2.31	526.58

Figure 6.17 shows the difference between true and augmented predictions. For the different algorithms, the progression appears to follow the same curve; however, in some cases, it is predicted below or above the expected line. The variation in the run-to-run failure cases may contribute to

353
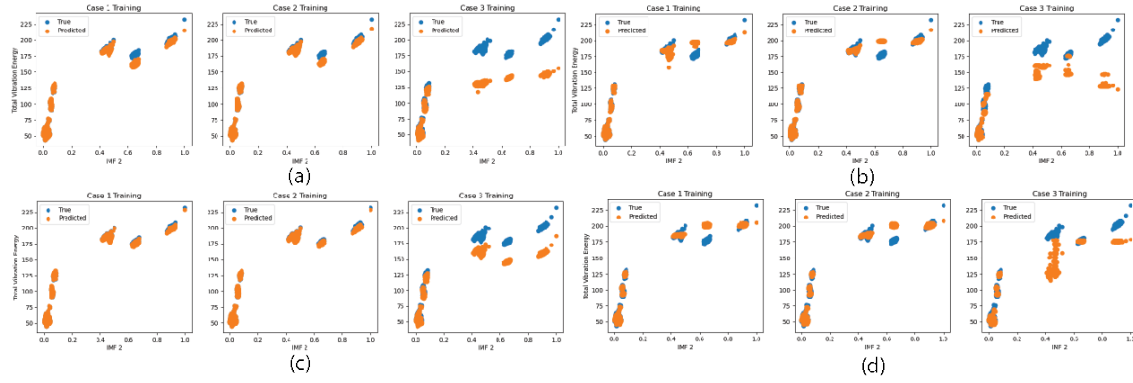
Figure 6.17: Regressors for select algorithms: (a) SVM, (b) RF, (c) NB, and (d) kNN

this variance. It is possible that the other features from Figure 6.15 cause variation in the algorithm response, similar to what occurred in the classifiers. The inclusion of the augmented cases appeared to reinforce the predictions for the true case and augmented case.

In this instance, the augmented data appears to help support the prediction of the failure condition; however, incorporating only augmented failure data does not appear to provide an accurate prediction for the bearing health. It is possible that the different regressors would perform better under the removal of certain features similar to the classifier cases.

6.5.1.3 LSTM Results

Table 6.27 contains the mean squared error for the LSTM models based on the different model configurations. Figure 6.18 shows the comparison of the regressor scores between the predicted and true cases. Based on the scores across the different training cases, the bearing appears to accurately determine the vibration content based on the defect case relative to the baseline case. The augmented cases appear to be able to better approximate the true failure data when considering only the augmented data. It is possible that the algorithm performance increase is a result of the LSTM ability to retain information relative to the different defect cases as the algorithm is training.

Table 6.27: Mean squared error for the LSTM models training scenarios

Model	Case 1	Case 2	Case 3
LSTM (6205 Configuration)	662.85	606.70	815.96
LSTM (6206 Configuration)	379.33	173.66	342.26
LSTM (6207 Configuration)	120.06	84.14	245.32

In addition, the algorithm dispersion appears to cover all of the different possible cases

354

relative to the bearing defects based on the results from Figure 6.18. Even with the increased error, there is complete coverage of the true cases with the predicted cases. Hence, the retention of information appears to support and verify the possibility of variance that could occur within the different defect cases. Hence, the possible predictive case is more robust with the inclusion of the augmented data versus leaving the data out.

6.5.2 Augmented Dataset: Spectral Augmentation

The 1687 RPM speed case was the data used from the 6205 case to provide an example of data augmentation using spectral augmentation (SA). Table 6.28 contains the makeup of the true and augmented data for the different healthy and failure stages. The "#K" describes the bearing designation for the different bearings, where the preceding percentage describes the amount of data used from that original data source. The total amount of bearing data used for the healthy case is split equally into test and training data. The spectral augmented healthy data is created using an average frequency response for the bearing baseline condition and the individual frequency responses for each bearing from the healthy data using the True healthy training data. The spectral augmented data is split into test and train sets based on a 50% split to the healthy data. The contrast to the physical scaling process evaluation is that this process assesses the creation of bearing data for bearings of similar size based on the modification of only their frequency content.

For the failure data, the true data are split into test and train data. For the failure data, an average frequency response is created for each failure condition based on the FFT results from the training case. The spectral failure data for each failure condition is created based on the difference between the different average frequency responses and the average baseline response. The spectral failure deltas are applied to the SA baseline data train data to create the SA failure conditions data. An example designation for the spectral augmented case is "SA Train Fatigue I.R. 1 mm (50% 3K, 4K, 5K, 6K SA Baseline Train)". The designation means that these data are represent the spectral augmented data for the 1 mm case and originally contains bearing data from the SA baseline case for bearings 3K, 4K, 5K, 6K, and 7K prior to adding in the spectral failure delta for the 1 mm case. The contamination case is similar except it is labeled as "Cont. Code #", which represents contamination and the contamination amount in the bearing.

The different training cases are then split into the following groups based on the data breakdown above. Each case has a training case and a test case. Case 1 contains only true data for the

355
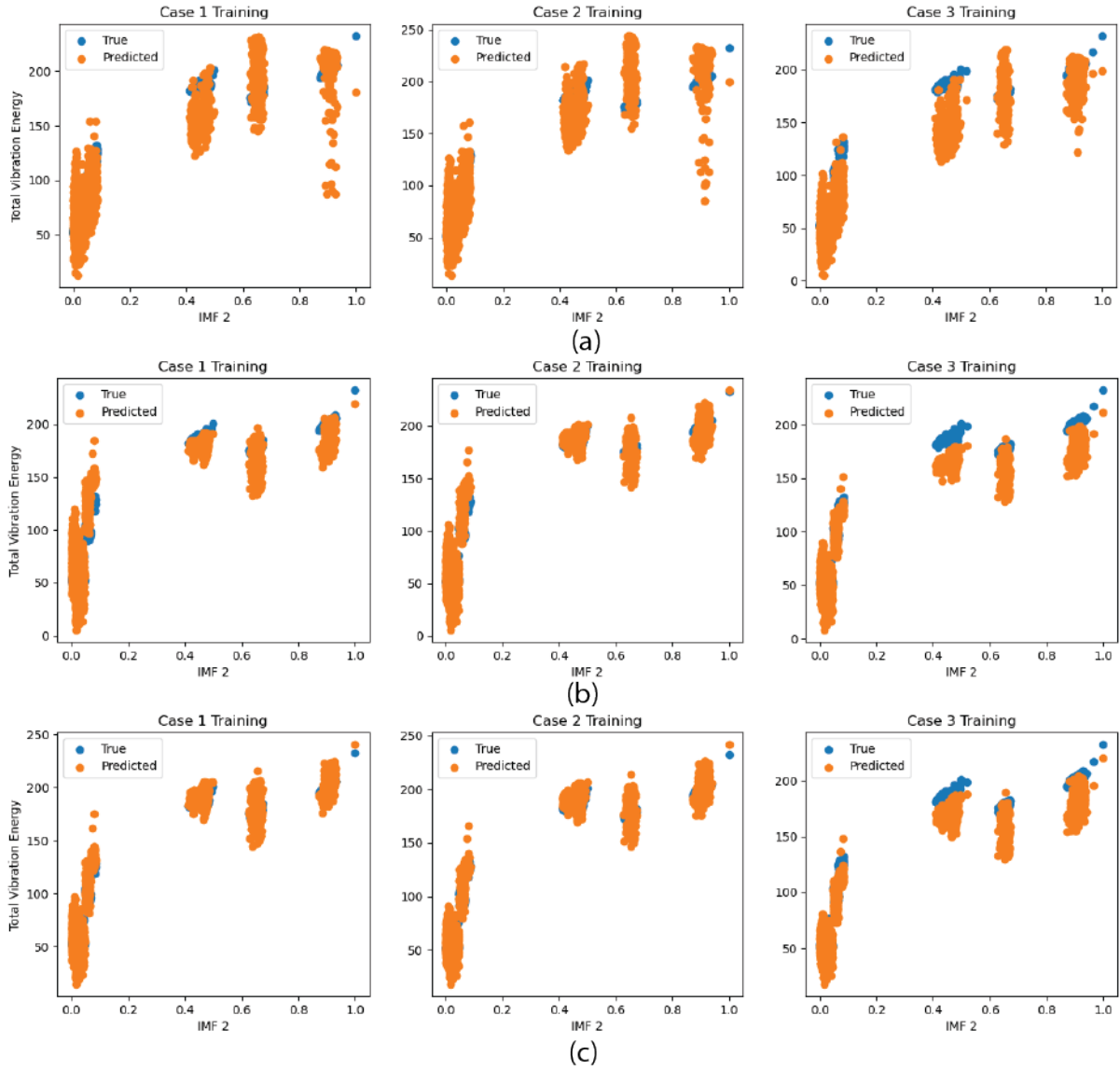
Figure 6.18: LSTM results for the different training cases using the (a) LSTM trained from the 6205 original data, (b) LSTM trained from the 6206 original data, and (c) LSTM trained from the 6207 original data

356

Table 6.28: Breakdown of data forming the augmented and true cases for testing in Section 6.5.2 for the 6205 bearing size. SA means spectral augmentation.

Data Types	True	Augmented
Healthy	Train Baseline (50% 3K, 50% 4K, 50% 5K, 50% 6K), Test Baseline (50% 3K, 50% 4K, 50% 5K, 50% 6K)	SA Baseline Train (50% Train, 50% 3K Train, 50% 4K Train, 50% 5K Train, 50% 6K Train), SA Baseline Test (50% Train, 50% 3K Train, 50% 4K Train, 50% 5K Train, 50% 6K Train)
Failure	Train Fatigue I.R. 1 mm (50% 12K), Train Fatigue I.R. 3 mm (50% 12K), Train Fatigue I.R. 7 mm (50% 12K), Train Cont. Code 20 (50% 12K), Train Cont. Code 24 (50% 5K), Train Cont. Code 25 (50% 5K) Train Cont. Code 0 (50% 5K), Test Fatigue I.R. 1 mm (50% 12K), Test Fatigue I.R. 3 mm (50% 12K), Test Fatigue I.R. 7 mm (50% 12K), Test Cont. Code 20 (50% 12K), Test Cont. Code 24 (50% 5K), Test Cont. Code 25 (50% 5K) Test Cont. Code 0 (50% 5K)	SA Train Fatigue I.R. 1 mm (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Train Fatigue I.R. 3 mm (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Train Fatigue I.R. 7 mm (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Train Cont. Code 20 (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Train Cont. Code 3K, 4K, 5K, 6K (50% 5K SA Baseline Train), SA Train Cont. Code 25 (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Train Cont. Code 0 (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Fatigue I.R. 1 mm (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Fatigue I.R. 3 mm (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Fatigue I.R. 7 mm (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Cont. Code 20 (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Cont. Code 24 (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Cont. Code 25 (50% 3K, 4K, 5K, 6K SA Baseline Train), SA Test Cont. Code 0 (50% 3K, 4K, 5K, 6K SA Baseline Train),

training case using only the "True Baseline Train", "True Fatigue Train", and "True Contamination Train" cases. The Case 1 test group contains the test data from the "True Baseline Test", "True Fatigue Test", "True Contamination Test","SA Baseline Train", "SA Baseline Test", "SA Fatigue Test", "SA Contamination Test", "SA Fatigue Train", and "SA Contamination Train". For the Case 2 training group, equal amounts of true and augmented bearing data are used from the "True Baseline Train", "True Fatigue Train", "True Contamination Train", "SA Baseline Train", "SA

357

Fatigue Train", and "SA Contamination Train". The augmented data are scaled down to match the true data size. The leftover data from the augmented train cases are incorporated into the test cases, which utilizes the test case counterparts from the training case. Case 3 represents the opposite of Case 1, where the true data cases in the training data are swapped for the counterpart augmented cases. The test case utilizes all of the remaining data for testing the algorithm. It should be noted that in this case the comparison stems from equal using equal percentages of bearing data based on the different training cases.

6.5.2.1   Classification Results

Table 6.29 contains the F1 scores for each algorithm used across the different training cases considered. The classifier performance was conducted over the same progression for both the contamination and fatigue cases. For the contamination cases, the spectral augmentation data performed fairly well. The probable reason for the high performance is the separability between the different classes. The DT classifier did not perform as well, but that could stem from possible overfitting with the bearing training dataset. As was noted in the data characterization from Chapter 4, the introduction of contamination into the data marked a clear difference between the baseline and contamination cases.

In contrast, though, the fatigue cases did not perform as well when considering the different training cases between the augmented and real data. The close similarities between the different defect stages could have caused an increased number of misclassifications that the inclusion of the augmented data would have hampered. Incidentally, the real data had a hard time approximating the augmented data as a result and did not perform as well until a necessary subset of the data was included from the analysis. The DT and NB classifiers performed poorly overall, which could stem from the DT classifier overfitting the data used for training. The NB classifier performance could have stemmed from the lack of independence between the features.

The classifier results perform well for determining the bearing diagnosis of different stages. However, the best case appears when the training case incorporates only the data from the augmented case. The reason for that increase in performance could stem from the variability in the baseline data from the initial testing. That variance is retained as part of the process. Hence, the reason for the augmented data performing "better" during the training process stems from the ability to approximate that true case, which is represented in the variance. However, the opposite is not

Table 6.29: F1 score for the failure progressions in contamination and fatigue scenarios based on the different training scenarios (Case 1/Case 2/Case 3)

Model	Contamination	Fatigue
Naieve Bayes (NB)	0.898 / 0.996 / 0.978	0.493 / 0.698 / 0.999
Support Vector Machine (SVM)	0.999 / 0.999 / 0.999	0.813 / 0.903 / 0.998
kNN	0.999 / 0.999 / 0.999	0.797 / 0.898 / 0.999
Decision Trees (DT)	0.990 / 0.999 / 0.781	0.677 / 0.513 / 0.999
Random Forest (RF)	0.999 / 0.999 / 0.996	0.751 / 0.819 / 0.999
Multilayer Perceptron (MLP)	0.999 / 0.999 / 0.999	0.794 / 0.918 / 0.999
Deep Multilayer Perceptron (DMLP)	0.999 / 0.999 / 0.999	0.912 / 0.898 / 0.999

necessarily true, especially for those data that are outliers in the baseline response. That way, there are elements from both true and augmented data cases to provide examples during the training process.

6.5.2.2 Regression Results

Table 6.30 contains the mean squared error for the regression algorithms tested on the three training cases. The error appeared to be greatest in the training Case 1 instance, similar to what occurred with the contamination algorithms. In this instance, the DT and RF regression algorithms perform the worst, even over the NB case. It is possible that the poorer performance could stem from overfitting in the training cases as opposed to the test case. The error is greatly reduced for both algorithms as augmented data is added to the training case. The Bayes regression and Multilayer Perceptron performed the best for all instances. It is possible that there is a discernible trend in the data that is easily approximated by the Bayes approach. In predicting the total vibration content of the signal, generally, it is expected that this increases with the different feature content. However, as is evident from the classifier approach, it is not necessarily a linear trend.

Table 6.30: Mean squared error for the different regression training scenarios considering both failure modes

Model	Case 1	Case 2	Case 3
Bayes Regression	280.45	17.11	93.52
Support Vector Machine (SVM)	322.6	19.00	112.19
kNN	467.33	150.19	172.50
Decision Trees (DT)	1365.89	313.40	133.16
Random Forest (RF)	3437.68	740.62	117.94
Multilayer Perceptron (MLP)	323.30	17.28	93.58
Deep Multilayer Perceptron (DMLP)	280.50	17.10	93.53

Figure 6.19 contains examples of different comparisons from the regression algorithms in the

(a) Bayes Regression case, (b) DT case, and (c) DMLP case with respect to the IMF 2 features. For the NB case, the predicted values appear to perform closely to the true values in each case, especially in Case 2 and 3. There are some errors associated in Case 1, but that could stem from possibility in similar target values that are then found over different feature values. The variation in the DT classifier is noticeable as a result and appears to stem primarily from overfitting in the augmented versus real data. As augmented data are added to the training case, the error for the different type of algorithms decreases. Even in the random forest instance, a high degree of overfitting was noted based on the algorithm perforamnce that was reduced by the introduction of the data variance. For the DMLP case, the results are similar to NB case showing that the predicted regression values are similar despite the differences in the algorithm structure. Hence, it is primarily the training data configuration that causes the variation in the algorithm results.

For the regression case, the algorithm scores did not perform as well in using only the real training data as opposed to the augmented data. The larger error could stem from two different aspects. The first consideration is the approximation of the real training data on the augmented data case. As was noted in the algorithm performance case, the bearing analysis did not perform as well when the augmented data was considered in the test case due to the variance retained in the baseline data. The second consideration stems from the amount of the data used in the three different cases. For the original training data case (Case 1) and the augmented case (Case 3), the amount of data used in training remained the same. The Case 2 case utilized double the data amount, due to the consideration of equal amounts for augmented and true training data.

6.5.2.3   LSTM Results

Table 6.31 contains the LSTM results for the different bearing cases using the 6205, 6206, and 6207 pretrained models for the LSTM cases. The a high mean squared error value means there was a greater difference between the predicted and expected (true) results. The mean squared error case appears much greater as compared to the validation cases. The LSTM results are much greater than what was noticed in the validation cases from the original data cases. The primary differences between the different algorithms appears to be the number of nodes used in the 6205 case and the 6206 and 6207 cases. In this instance, the primary difference is the number of nodes used to approximate the data between the 6205 case and the 6206 and 6207 cases. In this instance, however, the difference in the 6205 and 6207 cases is due to the different weights used in the LSTM models.
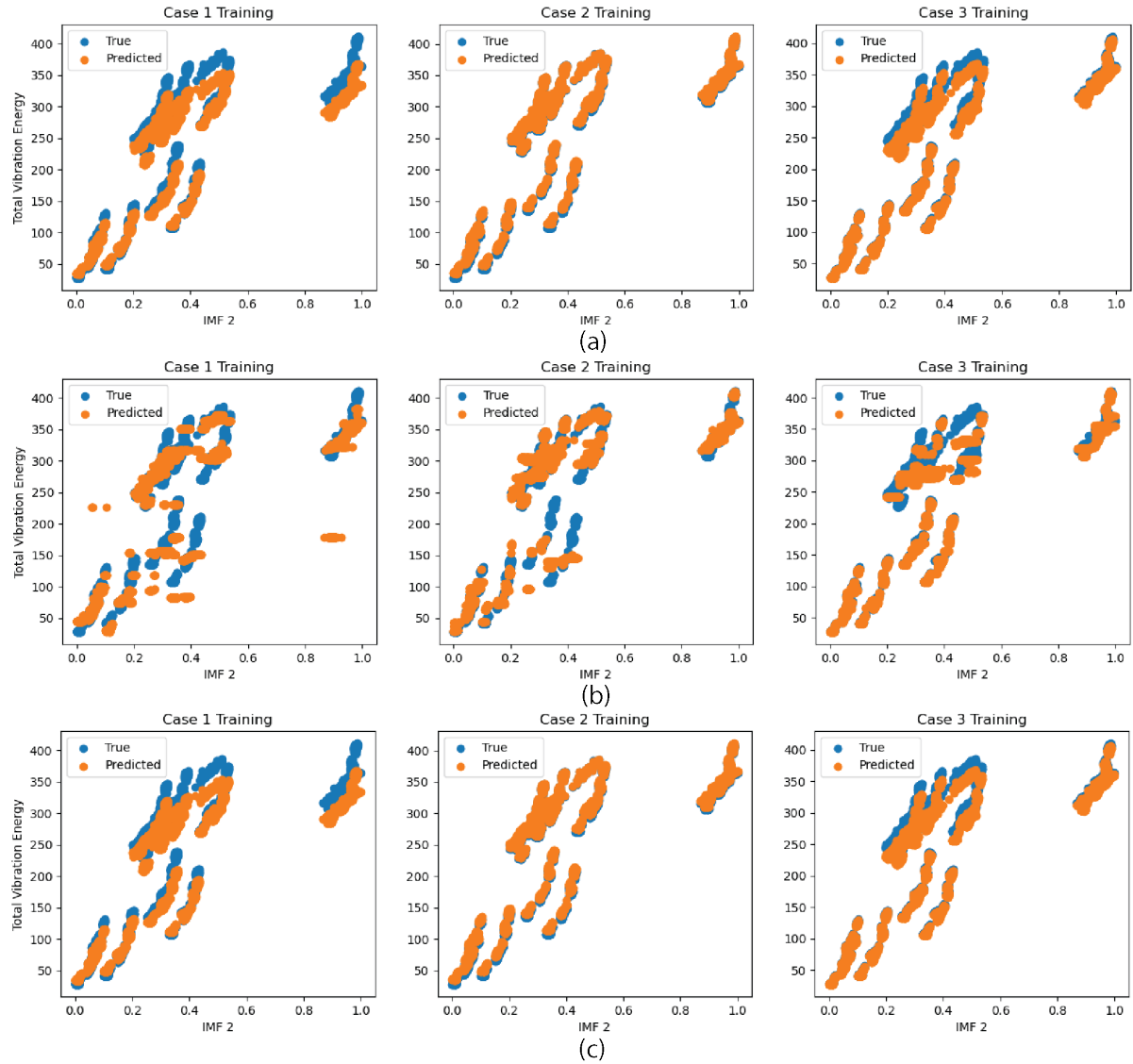
Figure 6.19: Regression results for the (a) NB case, (b) DT case, and (c) DMLP case

The difference in weights could stem from the number of epochs used during the training process. The difference in the model results could occur due to the poor approximation that the true data can cover the augmented data. This observation represents a difference as compared to the regression results and may stem from the data abilities to approximate the non-linear tendencies in the bearing data.

Figure 6.20 contains the LSTM results for the different model configurations trained on the original 6205, 6206, and 6207 data. The representative error values appear to conform to the model

361

Table 6.31: Mean squared error for the LSTM models training scenarios

Model	Case 1	Case 2	Case 3
LSTM (6205 Configuration)	1578.8	559.65	216.61
LSTM (6206 Configuration)	849.88	224.71	122.02
LSTM (6207 Configuration)	403.66	48.28	86.27

expectations. The Case 1 in the LSTM 6205 model appears to follow a similar trend, but does not fit the data well relative to the expected true values. It is possible that a longer training time (varying the epochs and original data) is necessary to provide better results to the algorithm. Even in the instance of the other case 1 algorithms, the fit does not appear to conform well with the other bearing data. The more likely scenario is that the bearing data does not approximate unseen augmented values well; that is the primary difference to the original data validation cases. In this instance, the augmented training data appears to approximate the true data values cases better than the true training data due to the coverage seen across the true and predicted cases. It is possible that this stems from incorporating more samples in the training case for the spectral augmented data versus the true bearing data.

For the LSTM results, the algorithm performs better when examples from both true and augmented data cases are considered in the algorithm training. However, the 6207 case appeared to deviate from the expected norm due to the difference in nodes and weights for the 6205 and 6206 algorithms. Hence, further training of the 6205 and 6206 cases could yield similar results as to the 6207 case. It should be noted that the Case 3 based on Figure 6.20 appeared to show the best convergence of results despite the Case 2 error having a better error value on average.

6.5.3 Augmentation Methods: Conclusions

Based on the algorithm results, the augmentation data performed well in different cases as compared to the bearing data cases. In this instance, the PS process was tested to create bearing data for a different bearing size based on the difference in the RMS values and the model expectations. From the PS process, the bearing data did fit well in terms of supplementing the training case (Case 2). However, it did not perform as well as replacing the true data case as shown in the Training Case 3. The observation could stem from the variation observed in the bearing data observed in the real training case versus the augmented bearing case. The PS data did fall within the expected range of the bearing data features as seen in Figure 6.15. However, it also shows that the variation
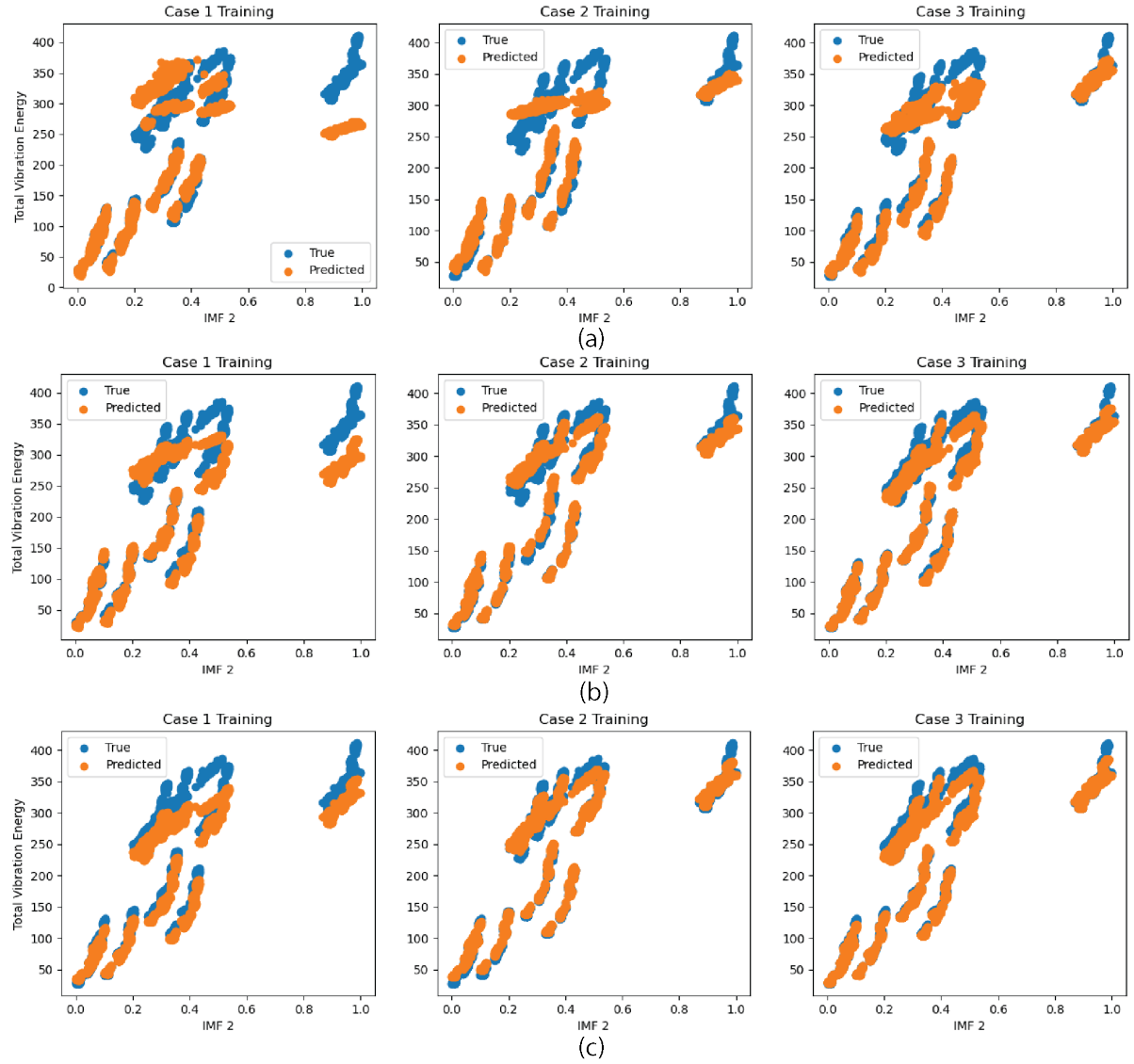
Figure 6.20: LSTM results for the different training cases using the (a) LSTM trained from the 6205 original data, (b) LSTM trained from the 6206 original data, and (c) LSTM trained from the 6207 original data

shown in the real data that was not accounted for in the augmented data.

For the spectral augmentation process, the algorithms performed poorly in the real data case as opposed to the augmented data case for the fatigue case. In this instance, the bearing data variance was found more so in the baseline case. The variance was retained as the transformed occurred using the spectral augmentation process. Hence, it is more apparent that whichever feature case, real versus augmented, the subsequent algorithm performance performs better when using

363

the training case with more variation to predict the different diagnosis method. This observation could explain why the PS process performed well in one particular instance, whereas the spectral augmentation preformed in a different manner.

As far as choosing the choosing the better algorithm performance, the PS process performed well in scaling data from different applications. However, if there are differences in the frequency peak locations between operating and failure cases, the bearing algorithms for the PS scaling did not perform as well. In the spectral augmentation process, the bearing data was able to retain the variation observed in the bearing data. The spectral subtraction incorporation helps to incorporate the potential differences within the bearing data not readily observed by the baseline data. However, the scaling aspect does not perform well between bearings of different sizes, due to the potential variance observed in the different bearing processes. In the spirit of the question, the spectral subtraction does perform better based on the results from the experiments using Case 3. However, it should be noted that the variability in the training data could influence those results.

6.6   Model Performance: Data Tuning

After seeing the effects on a mixture of true and augmented data, further refinement was considered to determine the minimum amount of training data needed for the different classifiers and regression cases and if there was an ideal quantity of real and augmented data to train the different systems. The ideal training and ideal ratio of augmented versus true training data are based on the different training cases considered in the augmented case. The ideal case is evaluated based on this scenario's minimum F1 score and minimum mean square error. The transfer method used to create the augmented data was the spectral augmentation method to consider both the contamination and fatigue cases for the 6205 bearings. Similarly, three different training cases were considered for the algorithms, where Case 1 considered algorithm training with only real data, Case 2 considered both real and augmented data, and Case 3 considered only augmented data. For the classifiers, the F1-score assessed the different test cases. mean-squared error was utilized to assess the different regressors

6.6.1 Ideal Training Amount

Figure 6.21 provides the difference in the algorithm results for (a) Case 1, (b) Case 2, and (c) Case 3. The x-axis characterizes the amount of training data, as a percentage, used relative to the different training cases. Hence for Figure 6.21 (a), the percentage of training data used characterizes the variation of training dat from the real features. From Case 1, the contamination classifiers did not require an extensive amount of training data and needed only approximately 10% of data to cross the 0.9 mark for the different machine learning algorithms. However, the DT classifiers appeared to experience a variation in the algorithm results as the training data amount. It is possible that this stemmed from variance in the training data. In the case of the NB classifier, the algorithm performance dipped as the amount of training data increased. It is possible this could stem from overfitting in the different algorithms as a result of the training process. In the fatigue performance, the algorithms never reached above 0.90, except in one instance for the DT classifier and one instance for the MLP classifier. The better performance in the contamination case versus the fatigue case could stem from more separability in the different training cases. For the regressor algorithm, the algorithm score did not appear to change significantly, except in the MLP classifier. Hence, based on the spectral augmentation process, the real training data does not provide an accurate approximation of the training data.

For the Case 2 algorithms (Figure 6.21 (b)), augmented data were utilized as part of the training process. In this instance, an equal amount of real and training data were used for the comparison and varied as a percentage of the combination of the two algorithms. For the contamination algorithms, the F1 score passes the 0.9 mark, at approximately 10% of the data used for each case. There is variation still present in the data as a result for the DT case. However, the algorithm performance does not change significantly after 45% of the training data used. In the fatigue case, the algorithm performance improves quicker over time; however, it still does not yet reach the 0.9 mark relative to the different training categories. The NB classifier performs steady despite the different training cases. The KNN and SVM also reach a plateau as the amount of training data increases at approximately 40%. For the regression analysis, there is less apparent variation observed as a result in the regression analysis. In the case of even the MLP classifier, the reduction of error occurs more rapidly for the different algorithms as is evident for the different cases. There is one anomaly that occurs with the DMLP classifier. That could stem from a possible variance in the training data that
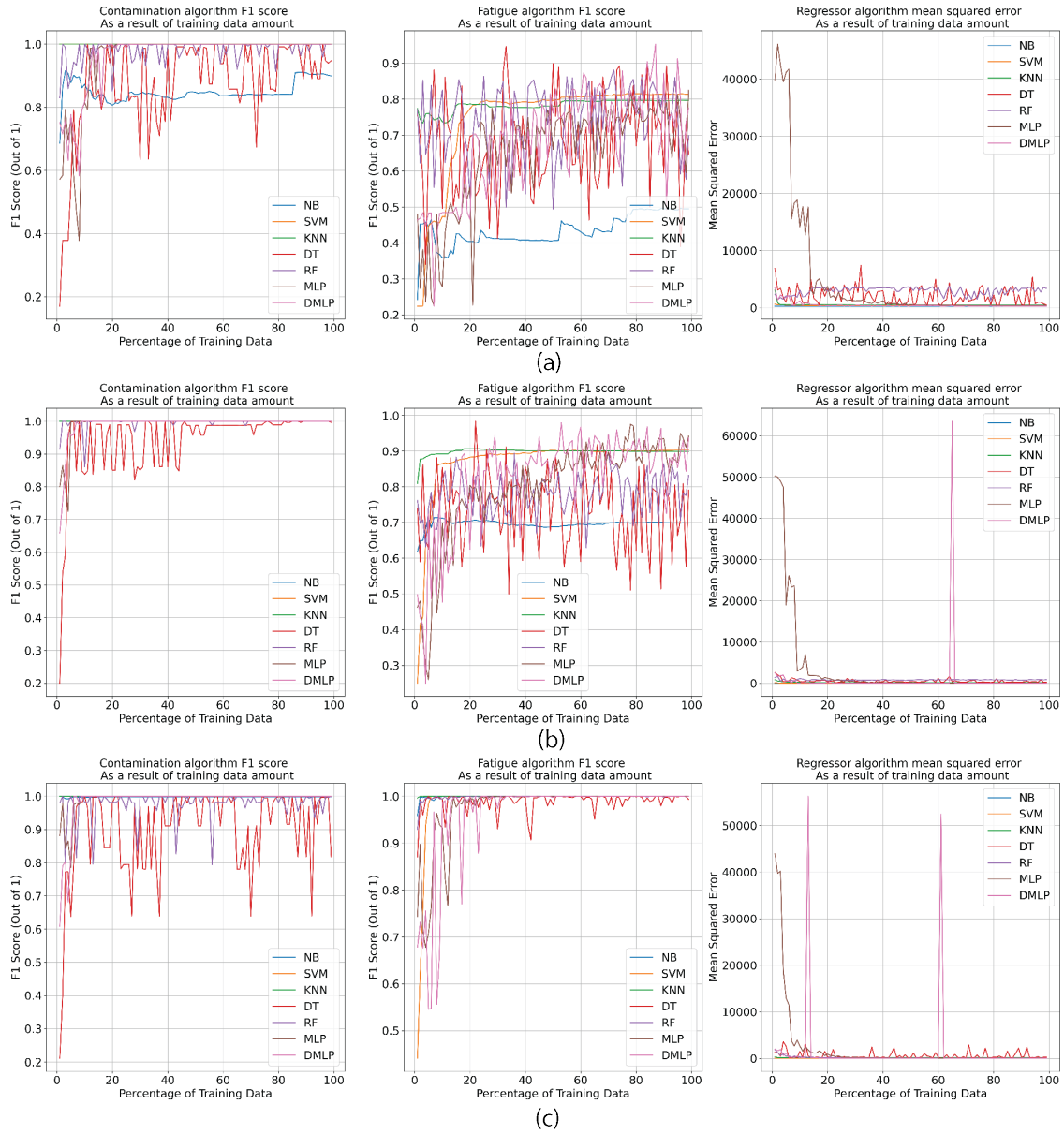
Figure 6.21: Classifier contamination performance, classifier fatigue performance, and regressor performance for the different training cases: (a) Case 1, (b) Case 2, (c) Case 3

was not anticipated. However, the overall error is reduced as the training data increases.

For the Case 3 performance, the algorithm error was further reduced across the three different categories. For the contamination case, the algorithms still performed at greater than the 0.9 mark once the data had reached an 10% of the expected training data. There were notable instances of variation from the DT classifier that could have stemmed from potential variance in the training

data. For the fatigue cases, the percent training data appears to perform closer to the 20% mark to reach a 0.9 or greater F1 score for the MLP, DMLP, and RF classifiers. For the regression analysis in the Case 3 training data, overall performance appears to be lower than in (a) Case 1 and (b) Case 2 regression results. There are still unexpected spikes in the data hinting at possible variation within the training data sets.

The DT classifier still does experience instances of overfitting through the different contamination, fatigue, and regression algorithms hinting at the variance in the data. It is possible that the greater influx of augmented data, as opposed to the real data, influenced the algorithm results as well. However, even with the influx of additional failure training data, the variation experienced in the different algorithms does appear to support that the variation of the augmented training data improves algorithm results more solely the amount of data in either case. If the results were to be normalized based on the data amount, the augmented training data would still reach a higher algorithm performance than the real bearing data. For determining an ideal training amount, though, it appears that the training data variance influences the results more so than whether the data comes from the real versus augmented domain.

As mentioned earlier, with the Physical Scaling method, there needed to be more failure data and knowledge to provide a prediction for the generated bearing data. The ideal amount of training data investigated with spectral augmentation highlights how it is possible to create additional failure data using physics-informed methods even with a small amount of failure data available. It could be possible to incorporate additional data types using the physics scaling method to provide a prediction for the expected bearing amplitude to further scale bearing data operating at different parameters to the system. As long as there is some basis in the training data related to the physical system source, low amounts of failure data are needed overall. It should be noted that the percentages of the different training cases tested were considered based on the respective datasets.

6.6.2   Augmented vs. True: Ideal Quantity

Whereas the prior section (Section 6.6.1) contains the overall assessment of the needed training data under the three different combination of training data, this section considers only the ratio of real training data versus augmented training data and the subsequent algorithm performance. Figure 6.22 considers the different algorithms and failure modes as a result of changing the amount of available data between the different cases. In this instance, the x-axis refers to the percentage

of real data used in the training approach versus the amount of augmented data. For example, the 99% along the x-axis corresponds to the case of when the training case holds approximately 99% of the real data and only 1% of the augmented data. Generally, based on the algorithm results, there is not a change in the algorithm performance as the type of data is varied between the different cases.
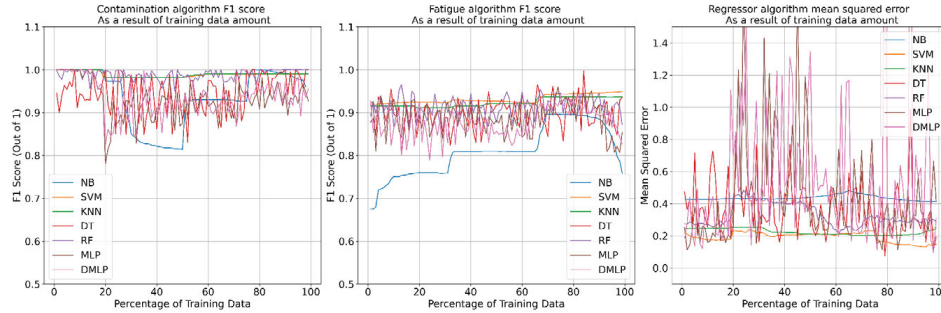
Figure 6.22: Bearing contamination and fatigue classifier and regressor performance as a result of the varying amounts of training data from the true data case represented as a whole of the training data

For the contamination and fatigue case, the NB classifier appeared to have the largest issue in the approximation of the data based on the ratio of the true versus augmented data. It was interesting to note that the NB classifier f1 score appeared to increase concerning an increase from the real training data. The other algorithms appeared not to be influenced by either combination of real versus augmented data. There is a high amount of variation seen between the two different cases as a result of the training process, suggesting that there is still a degree of unresolved variation in the training data as a result of the process. Incidentally, the true data could be causing the lower accuracy rate as the variation in the data amount from Case 3 does not necessarily appear in the training data despite the lower amount of training data used. The total amount of data used is equivalent more so with training Case 1, which has a much lower metric score as opposed to this training case.

Based on this performance, the augmented data could perform well over the different classifiers and regression cases. There is still some variability in the results, but the overall classifier and regressor performance remains high for the augmentation. The results support using the spectral augmentation method to create bearing data to resemble true failure modes from a physical system using scaled bearing data. Another version could incorporate the physical scaling process to support the transference of the spectral failure delta from similar applications.

6.7   Model Performance: Feature Windows

Data features are created by creating subset groups of features to obtain different results. These features windows are created statically or dynamically. Static feature windows means that the features do not change during the data preprocessing. Dynamic feature windows implies that the data groupings change as a result of different influences from the data. Kong *et al.* [549] created an adaptive time window that changed with respect the bearing condition. Since it is known that the bearing life changes exponentially, the bearing time windows change to reflect that exponential change and capture the relative bearing data.

In this essence, the feature window is explored to see how much the information transfer is affected by the change in the feature window size. The tuning of the window size would affect the amount of data needed to collect and what is transferred between different bearing systems. For example, reducing a text file of with 10 values in the format of ("##.####") to 1 value reduces the data file size from 88 bytes to 7 bytes. If it were to come from 40 values to one value, the size changes from 358 bytes to 7 bytes. On a continuous time collection scale, that is saving 1 MB approximately every three collections. The difference further scales when considering that vibration data requires raw continuous samples for frequency and statistical analysis and the data is collected on a scale of tens of thousands of samples.

Figure 6.23 shows the difference in the frequency representation due to changing the number of samples that make up the FFT under a constant sampling rate. The change in the number of samples changes the frequency resolution. As mentioned earlier, frequency resolution determines the number of bins and relative bin location along a frequency. The distribution of bins is equidistant based on the number of samples. Hence, If the data was collected at 100 Hz and the samples used in the FFT was 50, then there would a frequency bin every 2 Hz. If more samples were used than the sampling rate, say 200 samples collected at 100 Hz, then the frequency bin is every 0.5 Hz. Generally, the FFT resolution is configured based on the amount of data needed for sampling and the necessary granularity of signals. The sampling rate is generally set by the Nyquist criterion, which determines the maximum observable peak for the data. From there, the resolution of the FFT can be varied to determine the different sizes necessary for maximizing the information and data needed between different samples. Generally, a higher resolution is desired to focus on specific frequencies, some times at the first decimal point and lower. In these cases, the sample lengths

369
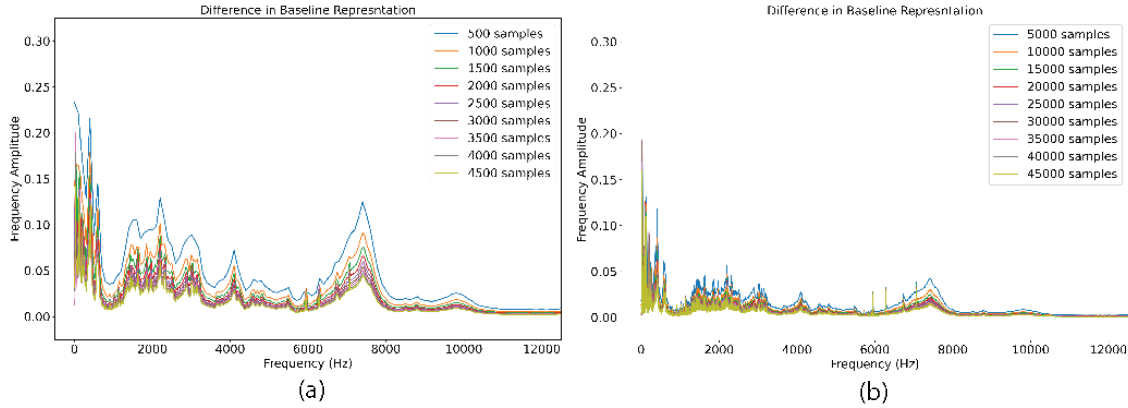
Figure 6.23: Change in average FFT when considering sample subgroups of 500s when the data are sampled at 50 kS/s

would need to be greater than the sampling rate as earlier illustrated to ensure that the bearing frequency is identifiable.

In this application, the bearing characteristic frequencies are critical for observation. However, they are grouped in to the IMF frequency energy bands for their consideration. Hence, it may be possible to reduce the amount of data needed or processed to yield a similar prediction as is currently gathered. For this section, the FFT resolution is varied based on the number of samples within each feature group and how it affects the results for transferring data between different bearing applications. This application tests the spectral augmentation method by varying the sample groups to create the relevant failure deltas to transfer information. The initial consideration considers reducing the FFT resolution by varying the sample size less than the sampling rate. The sample amounts are varied based on the samples: 500, 2500, 5000, 15000, 25000, and 35000.

Table 6.32 shows the model performance for different damage progressions when the damage progression is varied using the spectral augmentation method. Based on the response, it does not appear to affect the contamination data. It is possible that the separation in the contamination data is to great to be of any consequence to the algorithm performance. For the bearing fatigue case, there is much more of a classification error that appears to occur, possibly due to similarities within the baseline case as seen in Figure 6.15. That particular figure was for the 6207 case, but it is possible that a similar trend occurs here based on the original training data analysis in Section 6.3. The bearing classification scores appears to stabilize at approximately 5000 samples for the contamination case and 35000 for the fatigue case. In the case of the fatigue data, this is actually an

improvement over the bearing response seen in the Case 1 training data results from the augmented data evaluation. The DT classifier has a lower performance as a opposed to the other classifiers for the fatigue case, where the performance crosses the 0.9 mark at the 35000 samples mark. In the NB case, the performance does cross the 0.9 mark at 5000 samples, but dips back down in to the upper 0.8 range.

Table 6.32: Classifier F1 scores for determining different damage progressions for Case 1 Training Case

| Number of Samples | 500 | 2500 | 5000 | 15000 | 25000 | 35000 |
Algorithm						
NB Contamination	0.972	0.999	0.991	0.999	0.999	0.999
SVM Contamination	0.939	0.999	0.999	0.999	0.999	0.999
KNN Contamination	0.894	0.998	0.999	0.999	0.999	0.999
DT Contamination	0.999	0.999	0.999	0.999	0.999	0.998
RF Contamination	0.999	0.999	0.999	0.999	0.999	0.999
MLP Contamination	0.996	0.994	0.999	0.999	0.999	0.999
DMLP Contamination	0.997	0.998	0.999	0.999	0.999	0.999
NB Fatigue	0.695	0.705	0.904	0.874	0.874	0.872
SVM Fatigue	0.683	0.813	0.898	0.981	0.995	0.995
KNN Fatigue	0.652	0.793	0.904	0.981	0.995	0.995
DT Fatigue	0.693	0.744	0.781	0.842	0.845	0.920
RF Fatigue	0.723	0.777	0.867	0.915	0.983	0.977
MLP Fatigue	0.767	0.875	0.908	0.981	0.992	0.994
DMLP Fatigue	0.759	0.767	0.893	0.975	0.966	0.993

In this instance, the classifier performance values are based on the varying amount of true training data to predict both the true and augmented data accurately. In both the contamination and fatigue case, the classifier performances appeared to perform well at well below the sampling rate. That could support reducing the sampling rate by at least half in the case of the fatigue case. For the contamination case, it is possible to go lower. Another impact is the reduced amount of training data needed to store for retraining the algorithm. Reducing the amount of data needed per collection by at least half would decrease processing times significantly and the corresponding costs of storage and processing.

Hence, to see if the trend continues, the augmented data were tested using training Case 3, which considered only the augmented training data for each algorithm. Table 6.33 contains the algorithm F1 scores for each classifier based on the failure mode performance. Interestingly, when considering the augmented failure data for the algorithm performance, the fatigue case performed better over the smaller feature windows than the contamination case. In this instance, the contamination algorithms performed well at the 25000 samples. It is possible that the increased variability

371

in the contamination case as a result of the spectral augmentation method would have lowered the algorithm's performance. Nevertheless, the performance does appear to support the reduction in data for the augmented data; however, the opposite failure modes are considered.

Table 6.33: Classifier F1 scores for determining different damage progressions for Case 3 Training Case

| Number of Samples | 500 | 2500 | 5000 | 15000 | 25000 | 35000 |
Algorithm						
NB Contamination	0.431	0.972	0.988	0.998	0.998	0.987
SVM Contamination	0.247	0.962	0.944	0.999	0.999	0.999
KNN Contamination	0.246	0.957	0.946	0.999	0.999	0.999
DT Contamination	0.432	0.649	0.779	0.739	0.926	0.944
RF Contamination	0.303	0.715	0.982	0.926	0.987	0.953
MLP Contamination	0.293	0.885	0.904	0.980	0.999	0.999
DMLP Contamination	0.294	0.868	0.937	0.958	0.999	0.974
NB Fatigue	0.577	0.976	0.989	0.999	0.999	0.999
SVM Fatigue	0.671	0.985	0.980	0.999	0.999	0.999
KNN Fatigue	0.705	0.981	0.980	0.999	0.999	0.999
DT Fatigue	0.621	0.947	0.953	0.953	0.988	0.903
RF Fatigue	0.630	0.972	0.983	0.999	0.999	0.999
MLP Fatigue	0.731	0.981	0.990	0.999	0.999	0.999
DMLP Fatigue	0.667	0.981	0.992	0.999	0.999	0.999

Based on the overview taken here, it is possible to observe that the feature window size does influence the algorithm F1 score in determining the equipment status. However, the influence of whether the equipment is healthy or not appears to occur based on whether the system contains either true or augmented data. The contamination data can have smaller feature windows for the true case and still accurately determine the equipment health. For the fatigue case, the augmented data case allowed for a reduced feature window size. The reduced data amount for the features allows for a reduction of the potential cost from storage, processing, and prediction while maintaining the same F1 score.

6.8   Data Evaluation Conclusions

The primary purpose of this chapter was to evaluate how well any generated data performed when used for training in classifier, regression, and deep learning algorithms. In addition to testing the model performance, several other model considerations were tested, such as the minimum training data amount and an ideal ratio of augmented to true features. For the data evaluation, the following questions were the primary focus of the data evaluation:

i.) Research Question 3.1: Is there an ideal amount of real versus augmented failure data for deploying offline data to a bearing diagnostic system?

ii.) Research Question 3.2: What augmented data process (data augmentation or physics scaling) increases the model performance regarding the assigned algorithm metrics?

iii.) Research Question 3.3: Does feature window optimization for bearing features improve bearing fault diagnosis regarding the assigned algorithm metrics?

In addition to these questions, the following questions were also considered to finish answering these questions

i.) Research Question 1.3: How much data are needed to train a bearing condition monitoring system for diverse failure modes?

ii.) Research Question 2.3: Are the physical scaling and spectral augmentation sufficient for translating data from the laboratory environment to the real world when considering similar applications?

The beginning portion of the analysis considered establishing a baseline expectation for how well the models performed based on the training data provided. A grid search was undertaken to identify the particular hyperparameters associated with each model. These models were then used interchangeably to identify if transfer learning was possible. Based on the model performance, the models for the individual bearing sizes were found to be interchangeable based on the algorithm performance. After the algorithm performance and model parameters were characterized, the augmentation data was tested based on mixtures with and without the true training data. The spectral augmentation case was primarily tested due to having performed better relative to the data transferability scenario.

For Research Question 1.3, an ideal amount of training data was found relative to the different training cases when including true and augmented data for the spectral augmentation case. The true failure cases require more bearing data to account for the variability introduced through the augmentation cases. The ideal amount of training data from both the augmented and true case based on the algorithm performance was approximately 10 - 20% based on the different failure modes and the algorithm chosen. For the substitution of augmented data versus true training data, the ideal amount of bearing data from the augmented failure cases needed was approximately 10% -

373

20%. The additional bearing data introduced a wider range of variability in the algorithm to account for both the true and augmentation cases.

For Research Question 2.3, the data transferability was originally assessed to translate from a laboratory environment to production equipment. However, the physical scaling method was tested by scaling bearing data from the 6205 to the 6207 case due to a production system's lack of failure data. In this case, the models performed poorly for the physical scaling process based on considering only the changes to the time domain value. The spectral augmentation process performed better based on the tested performance using either augmentation method. However, the spectral augmentation method relied on true failure data to provide part of the augmentation structure. The physical scaling process does not require representative data to create failure data related to the system. In one vein, it was possible to show that the combination of spectral augmentation and the physics scaling process could create failure data from different operating parameters (Section 5.3). In the spirit of the question, the best overall performance stemmed from the spectral augmentation method; however, combining both methods would allow for a better representation of true versus augmented data for both processes.

Returning to Research Questions 3.1 and 3.2, the ideal amount of real versus augmented training data did appear to influence the algorithm results based on the augmentation process for the fatigue case. For the three training cases, the PS results appeared to perform better for the experimental data (Case 1 training) than the augmented data (Case 3 training). In the SA results, the augmented training data performed better than the experimental data. On further evaluation of the results, it was found that the variance of the training data could influence the results greater than the use of real versus augmented data. The contamination data did not have as much overlap between the different data classes; hence, it was able to have higher-performing results. Hence, more than an ideal amount, the variance in the data was found as the influencing parameter. It was found that when using the spectral augmentation technique based on the training dataset, a higher performance for the contamination data was found between 10% (Case 1) and 20% Case 2, depending on the training data composition. For the fatigue case, only Case 3 ever reached above the 0.9 mark for the classifiers, and at least 20% of the augmented data was needed for that training.

For the data substitution between experimental and augmented, the bearing data algorithm performance does not appear to change significantly on average, except in the NB case. There is variation in the algorithm results for the DT, MLP, DMLP, and RF cases, but they are within

± 5% of each other. For Research Question 3.3, the feature window size was investigated but not necessarily optimized. Varying window sizes based on the number of samples per FFT were evaluated to determine if the classifier performance was diminished based on the window size selected. The results for the different classifiers for the true and augmented bearing data varied depending on the data type (true or augmented, failure versus contamination). However, the overall consensus was found to determine that based on the amount of failure data, at least half of the samples could be reduced, meaning that potentially, in a condition monitoring system, the processing time and storage size could be cut in half as a result.

Table 6.34 summarizes the current status of all research questions related to the work. At this point, satisfactory findings have been provided to all research questions. These are recapped in the conclusions section and describe future work to consider. For the work conducted in this chapter (Chapter 6), future work should consider the combination of the spectral failure deltas with the physics scaling process to enhance the spectral augmentation method further.

Formulation of a Digital Triple		
RA 1: Data Generation	RA 2: Data Transferability	RA 3: Data Analysis
RQ 1.1	RQ 2.1	RQ 3.1
RQ 1.2	RQ 2.2	RQ 3.2
RQ 1.3	RQ 2.3	RQ 3.3

Table 6.34: Current Status of the Research Goal as of Chapter 5

Chapter 7

Intellectual Merit

Currently, condition monitoring systems for equipment rely on two data sources: (a) generated test data of select specific applications and failure modes that may not relate to equipment in the industry and (b) real-world production applications where data quality is not verifiable. Establishing a surrogate system for the production system data pipeline is necessary to enable comprehensive failure profiles for robust predictive analytic training and retraining. This work seeks to understand and model the (a) generation of failure data on a surrogate system to build a comprehensive failure profile for real-world production equipment, (b) the transferability of failure data from a surrogate system to a production system, and the (c) the transferability of component surrogate failure data between multiple production systems with similar characteristics. The proposed approach enables integrating surrogate/laboratory data into the production environment with multi-sensor data for multiple failure modes, enabling scalability and the ability to deploy to multiple applications. The subsequent results would allow for the transition of designed condition monitoring systems, based on surrogate and production data, into the production environment without requiring failure data of the existing production system as consideration.

In the three areas related to condition monitoring, this research has proposed new science to address the lack of available data and the increased predictive power of condition monitoring systems. The Digital Triplet for Condition Monitoring was conceived to address these deficiencies. This research seeks to create methods for integrating the third triplet (Surrogate Triplet) into the condition monitoring system. Digital Twins are already an established tool; however, this research work provides the Surrogate Triplet as a means to take data from similar systems and utilize them

for training condition monitoring systems.

In Chapter 4, a purposeful failure method was established to generate different types of bearing data. General formulas had been proposed in the past related to design and ISO condition monitoring guidelines. The methodology (*Purposeful Failure Methodology* (PFailM)) is the first design method to propose creating a separate system to generate data for use in training condition monitoring systems for production equipment. The methodology was created considering ISO, design, and prior dataset characteristics to ensure that the system's design meshes with the existing knowledge of the physical systems and provides meaningful results. Using this methodology, a bearing dataset was generated. As far as this research work is concerned, this is the only bearing dataset that tests multiple failure modes under different progressions (4), under different test types (3), with multiple speed conditions (3), and with different loading conditions (2). Table 7.1 contains a breakdown of the dataset's characteristics covered in this research work.

Table 7.1: Breakdown of the bearing dataset covered in this research work

Failure Modes	Test Types	Speed Conditions	Loading Conditions	
Brinelling	Load-varying	1687 RPM	530 N	
Lubrication	Speed-varying	1103 RPM	614	(load-varying case)
Fatigue	Steady State	640 RPM		
Contamination				

The dataset created exhibits the proposed benefits of utilizing the Purposeful Failure Method. The collection methods are well-detailed, as well as the creation of the test system and how it resembles the expected production system. Different analysis techniques (Hilbert Huang Transform, Fast Fourier Transform, and statistical values) are employed to identify the critical differences in the bearing data due to the data generation. The dataset is the first to compare fatigue and contamination data with the same system, albeit with separate failure modes. However, the system establishment allows additional systems to integrate with the provided base and serve as a platform for new configurations. Furthermore, the data documentation offers a basis for the successive data transfer and analysis chapters. It should be noted that there is variability in the data, possibly due to the changes in the environmental and user conditions. These are noted in the data collection findings.

In Chapter 5, an alternative approach was proposed to address data imbalance and unavailability in condition monitoring for production systems by transferring data based on first-order

principles and expected phenomena with bearings. Physics scaling provides a new approach to relating data between different system configurations before training condition monitoring systems based on the expected vibration response of the bearing. While physics modeling of bearings has occurred to obtain frequency response, applying the response to scale the data has not been explored substantially. With the physical model, the experimental data could scale and retain their original frequency data to resemble similar characteristics with the physics model for the same operating parameters between different bearings. In another vein, a frequency stretch was shown to apply to the bearing to address differences in the operating conditions of particular bearing sizes. An alternative method known as spectral augmentation was established to create bearing data with a frequency similar to the other generated bearing data. The manipulation of the frequency in a direct manner represents a new and novel approach to bearing data augmentation. The data manipulation allows for the inclusion of frequency phenomena commonly associated with not commonly modeled failure responses, such as contamination. The application could serve as a means to increase the algorithm robustness to potential variation from

In Chapter 6, this research compares the augmented data versus collected experimental data to determine if there are ideal quantities of training data for different failure modes. A comparative analysis of classifiers, regressors, deep learning, and unsupervised learning techniques is used to define different patterns in the experimental data. The same configurations are used on different training cases of experimental and augmented data, where in observations are made possible based on the different combinations of the two data types. In this process, it is found that the data variance during training influences the model results more than the amount of training data, suggesting that further evaluation into accounting for data variability is required. The analysis also considered the interchangeability of real versus augmented data and commented on the results. In addition, the analysis considered the feature window tuning to minimize the amount of samples needed to create features.

Chapter 8

Broader Impacts

This work will significantly impact the integration of digital technologies for condition monitoring, not only for manufacturing equipment but also for power generation, transportation, and medical devices. The methodology for generating training data will enable standards of practice regarding data verification before training and implementing prediction systems in these industries. The CMDTri system will foster collaboration between industrial enterprises, which provide real-world system data, and laboratory environments, which would provide surrogate system data. A further benefit from this relationship is that the surrogate system for condition monitoring in a non-industrial environment enables training maintenance personnel to use digital technology for equipment repair decision-making before use in an industrial environment. Collorary to this benefit, the surrogate system would provide real-world scenarios to incorporate into class environments for teaching students the benefit of digital technology for anticipating failure or process events across multiple industries. The training and teaching from surrogate systems would allow for more understanding and purposeful integration of digital technologies across different sectors at all leadership and employee levels.

8.1 Financial Benefit for Condition Monitoring Digital Triplet

In truth, it is not simple to approximate the justification for a condition monitoring system based on a general cost. The final cost of failure is unique to each business and industry and largely depends on the maintenance strategy adopted. However, a general justification for a condition

monitoring system is when C_{cm} (cost of a condition monitoring system) is less than the potential cost from C_dt (cost of unexpected downtime). Hence, the return on investment is greater by utilizing the condition monitoring system to prevent equipment failure than allowing for the failure to occur. Therefore, the return on investment of preventing one failure and then successive failures justifies the reason for running and maintaining the system. However, when there are multiple possibilities for failures in the manufacturing system, multiple condition monitoring systems may be required to diagnose each possible failure. Hence, eventually, the cost of multiple condition monitoring systems may outweigh the benefit of unexpected downtime. The cost is further compounded by an additional time required to set up the integration of the condition monitoring system to the production line.

The primary benefit of introducing the digital triplet is to reduce the cost associated with training the system ($C_{training}$). The cost is broken down into the training cost for people to use the system and the actual cost to train the condition monitoring system with data. First, the people training cost is substantial, estimated at \$ 1,2000 per person in general from a 2014 industry report [550]. The cost to acquire a vibration training certification is approximately \$ 1,000+ based on offerings listed online[1]. This cost does not consider the additional cost needed to learn about the hardware and software required for an industrial system. Overall the individual cost to a person to familiarize a person with a condition monitoring system with vibration data could vary upwards of \$ 2,000, which may be unavoidable due to the technical necessities.

However, the further variation in the cost could stem from training the system if the data are not present, especially if failure modes are missing. It may take time to receive the required training data if considering data only from the production data. That data requires further verification to determine what occurred during the collection and assign appropriate labels. That could further increase training time and associated costs with the system integration. It is estimated that at least 80% of any machine learning project consists of data labeling and preparation before analysis [381]. Hence, in manpower hours and cost, that portion of collecting the necessary training data could become a full-time job in addition to performing other delegated activities in a manufacturing operation.

In addition, there is a cost to train and retrain the model. Li provided the framework of a tool to calculate the cost to retrain a model based on the computational power and time to perform the task [551]. In addition, there is a cost associated with storing the training data needed.

---

[1]Accessed from: https://www.vi-institute.org/correspondence-courses-self-study/ on 10/01/2023

Azure provides a price of \$ 0.16 to \$ 0.015 per gigabyte used[2]. It should be noted that we are only considering the data storage costs and not the ensuing analytics package. Using the cost calculator[3] with an analytics package, 1000 GB of reserved storage can cost around \$ 250. The amount of data needed to track one type of failure changes based on the situation, as shown with the different datasets in Table 4.7. In the essence of long-term collection as depicted in the IMS dataset [408], four bearings with only one sensor for each bearing over a month is about 4 GB, collecting 20 kS/s for one second every 15 minutes (based on test 3 size). A manufacturing facility has possibly thousands, if not hundreds of thousands of bearings. That also does not include the other type of data collection that could ensue based on specialized applications. If each bearing was under the same data collection plan, that could easily exceed the 1000 GB limit in a month.

Hence, a cost equation could be made in addition to the cost breakdown of failure provided earlier in Section 2.2.2. Equation 8.1 considers the total cost of training as a result of both the people skills required and the data costs. $C_{skillstraining}$ considers the necessary skills training to understand the basic concepts of condition monitoring and hardware training. The $C_{datastorage}$ considers the cost per gigabyte (other storage sizes acceptable) concerning the amount of data sampled (S_{data}) per collection (C_{data}) over some time (T_{data}) and for the determined amount of failure modes ($F_{failuremodes}$) that the system must consider. The model retraining cost considers the time and computational power needed to train the model after gathering the necessary data and choosing the proper algorithm ($C_{modelretraining}$). The data labeling costs ($C_{labeling}$) are needed to preprocess the data for the model training and to verify the model predictions.

$$C_{training} = C_{skillstraining} + C_{datastorage} * S_{data} * C_{data} * T_{data} * F_{failuremodes} + C_{modelretraining} + C_{labeling}$$
(8.1)

In a research project conducted with vibration sensors [552], monitoring seven sensors at a 1000 Hz sampling rate for six months yielded approximately 21 GB every two weeks in data storage. Based on the Azure storage cost, though, that would cause \$ 10.5 on a monthly accumulation. A couple of factors could increase that amount based on the frequency of collecting data. Increasing the amount of data collected (sampling rate) and the frequency of data collection would increase the amount of data stored. In addition, the failure point is theoretically unknown. Assumptions can be

---

[2] Accessed from: https://azure.microsoft.com/en-us/pricing/details/storage/files/ on 10/01/2023
[3] Accessed from: https://azure.microsoft.com/en-us/pricing/calculator/ on 10/01/2023

made based on industrial limits and equipment operation. However, these limits still carry uncertainty when applied to the data, and failure points may not be reached or missed when previously unknown.

The model training cost increases as the amount of data increases since the data increase requires more computational power for processing and time to train the model. In addition to the training time, the data labeling increases as the amount of data increases. The data labeling could eventually have an automatic system that provides labels for multiple equipment applications; however, the initial configuration could take time, depending on the user and application, to ensure low bias in the labeling response. Hence, there can be a significant amount of time between introducing a condition monitoring system and before predictions appear from the system.

Using the surrogate triplet within a condition monitoring system can reduce that training time and subsequent cost by coming pretrained based on known data related to the system. The difference between the DTr process and transfer learning is the application of known and explainable data versus only the trained model parameters. As was seen with Section 6.4, there was still some amount of error found with the transferred model from one bearing application to another bearing application on retrained data. As the digital system used in the monitoring ingests data from the surrogate system, theoretical adjustments could be made to the algorithm training process to tune the data to resemble more of the expected outcome (the physics scaling and spectral augmentation process) based on known observations related to the equipment. The increased data could improve the algorithm's robustness to possible variation from any deviations in the system as a result of past repairs or training from the operating condition.

The overall basis of incorporating the Digital Triplet is to reduce the need for data from the production environment, thereby increasing the maintenance performance with trained models easily deployed to the production environment. Reducing the implementation time and costs through the surrogate triplet helps to equalize the overall cost of the condition monitoring system. The potential benefit allows for scheduled maintenance to occur, thereby mitigating the possibility of hidden costs, such as production defects, environmental hazards, labor costs, or injury. Similar applications can stem from prototype testing from the design process, such as wind tunnel testing for different fluid properties or crash testing for vehicles. In turn, the barrier to entry related to cost is reduced for the overall system.

8.2 Trainig and Teaching

In addition to the general decrease in the condition monitoring cost and implementation time, the other aspect of using a Surrogate Triplet is the retention of knowledge related to condition monitoring systems. One of the aspects that occurs during an equipment failure is an investigation of how that failure happened and the possible steps to remedy that problem [6]. However, as that training occurs on the equipment in the production line, there is a greater emphasis on attempting to repair and restart the production line rather than understanding how the failure occurred and methods to prevent the loss in the future.

In the surrogate triplet, like during prototype testing, maintenance engineers can learn how equipment breaks in a controlled environment and how to interface with the condition monitoring equipment before the CM system deployment to the manufacturing floor. This training environment could increase the operation costs to maintain such an environment. However, the cost should be relatively small compared to not knowing how to react when a failure occurs in the production environment. Most manufacturing operations maintain a training cell for testing a new process. The surrogate triplet could be incorporated into a similar environment and adapted to match critical equipment on the manufacturing floor.

The retained and generated knowledge from the surrogate triplet would enable maintenance workers to practice reacting to potential impending alarms on the maintenance floor. Identifying and understanding the potential information coming from the condition monitoring system, coupled with knowledge of the potential failure, reduces the possibility of experiencing unexpected downtime on the manufacturing floor. As changes occur to the system, such as updates to the equipment, the DTr serves as a manner to track the changes and allow for the understanding of how any potential new equipment may affect the existing configuration of the DTr and the subsequent changes in the data.

Chapter 9

Conclusions & Future Work

9.1   Conclusions

In this research work, the primary goal was to create a new method for thinking about how data are created and applied to production equipment in manufacturing. To that end, the processes designed as part of the system were meant to ease the transfer of data, information, and knowledge through the system. The challenge tackled in this research is creating and comparing data between similar rotational equipment (bearings). This research discussed the following areas related to data generation, data transfer, and evaluating these data algorithms. To that end, research questions were identified to guide the discussions of new knowledge related to the bearing condition monitoring and the formulation of the Digital Triplet.

For the DTr, Chapter 4 considered the process for creating a surrogate triplet that complements the physical triplet and the data generation. During the formulation of the Digital Triplet, existing bearing datasets were rated based on how well they performed on certain dataset attributes, such as the appropriate amount of data, believability, and data verification. The wide dispersion of data generation techniques made it difficult to relate one dataset to another, coupled with missing contextual knowledge. Hence, the *Purposeful Failure Methodology* was utilized to create the experimental system and test plans for generating the data to ensure a complete understanding of knowledge related to the system. The data generation occurs based on the physical triplet failure modes and until a data amount by the user is reached. It was noted that there was a large amount of variability in the training data that could stem from untracked changes between users of PFailM.

The variability is not unexpected but is noted for documentation and reference in future work.

For the DTr, Chapter 5 considered the methods for relating the data to the physical triplet and utilizing the data for training machine learning algorithms. Prior methods have utilized scaling laws and dimensionless numbers for relating the vibration of different structures utilizing similitude. This approach used a physics model (Physics Scaling process) and a physics-informed method (Spectral Augmentation) to relate data between two systems based on the data generated from the surrogate triplet to the physical triplet. The spectral augmentation performed better based on utilizing existing data to inform the creation of new data. The physics scaling process did not require using existing data to inform and create bearing data. However, the frequency approximation did not perform well between the physical and surrogate triplets, leading to variation in the statistical significance results.

For the DTr, Chapter 6 considered evaluating the true and augmented data for how they performed in training condition monitoring systems. It was found that the variability noted in the data generation did appear to influence the algorithm results for determining the efficacy of augmented data versus real data. Hence, it was clear there were certain attributes between both data attributes that performed well. However, regarding the augmentation data performance, the spectral augmented data appeared to perform better based on comparing the fatigue cases. The amount of training data appeared to reach a threshold point after a certain amount of true and augmented data were used when implementing the spectral augmentation process based on the particular failure mode. Similarly, the spectral augmentation process did not appear to change the algorithm performance when changing the true versus augmented bearing data ratio for the algorithm training, except in select classifier cases (NB). These classifiers do have a known difficulty, though, concerning approximating variance in training data. The feature window size variation appeared to show some necessary information needed for retention in the augmentation case; however, a reduction in the necessary amount of data was observed, which could further point to a reduction in the necessary dataset size. Reducing the data size during collection would reduce the processing and analytic time for the production system.

The overall DTr was created to relate these three topics (data generation, data transferability, and data evaluation and classifier performance) to improve condition monitoring systems for bearing diagnostics and prognostics by creating data on an offline training system. Through the relationships and models tested, the data generated from an offline system can improve diagnostics

and prognostics for bearing systems, depending on the amount and variability within the training dataset. The overall benefit of the system stems from creating data with relationships established for application in another system. For component-sized systems, such as bearings, these relationships work based on the well-understood relationships of the component physics. The algorithm performance supports using the augmented data between similar systems for training based on frequency response considerations and similar time-series data between the surrogate and physical triplet. In deployment, the training of this system and subsequent retraining to account for possible variations reduces the overall necessary time for deploying condition monitoring systems to a production environment.

In this formulation of the Digital Triplet for condition monitoring, the surrogate triplet serves as the primary mode of creating and transferring data between different applications related to bearings. In conducting this research and providing answers to the questions listed above, a standardized pipeline for training data is created to ensure better algorithm performance that utilizes the physics related to the target system. The underlying physics is then applied to the data from the source system and scaled accordingly to resemble the output from the production system. This reduces the time and cost needed to train these condition monitoring systems as they are set up on the production line.

9.2   Future Work

Future work could consider five primary areas: the interface between people and condition monitoring systems, incorporating contamination modeling into the bearing physics scaling process, assessing the potential for combination failure modes, design for condition monitoring, and training data variance. The interface between computational systems and people is a diverse field that studies how condition monitoring systems could affect the decision-making process in manufacturing environments. In turn, system trust and the interpretation of information could help streamline the ease of adoption of these systems into other systems outside of manufacturing or relate new methods to manufacturing equipment. For bearings specifically, further consideration could investigate the incorporation of contamination modes and how to model the system wear as a result of the continuous impact of the particles. One possible method is the incorporation of many small defects from a bearing impact point of view with the lubrication models already derived to investigate lu-

brication starvation. Another item for future work is the investigation of combination failure modes. Combination failure modes are failure modes that comprise elements from two or more failures. For example, a combination failure mode could be a combination of lubrication starvation, contamination, and fatigue wear acting on the bearing. However, the severity of the prediction could change depending on which failure mode is the primary. Hence, further study is needed to investigate the identification of these failure modes and how to handle the defect progression as a result.

In addition, the work adds to the furtherance of design processes for condition monitoring, primarily those related to the data requirements. Currently, condition monitoring is usually seen as an add-on system. However, with this work, it is possible to consider incorporating these principles in condition monitoring as part of the design process. Another aspect is assessing the necessary amount of training data based on the variance in the dataset. Further work could consider the minimum data variance needed to train a condition monitoring system versus the quantity. As was seen in the research work, the variation or lack thereof in certain instances influenced the algorithm results for the different test cases. These future work items would further enhance condition monitoring systems by increasing the ease of use for these systems and investigating new potential sources of failure for rotational equipment. Furthermore, other equipment should be considered at the same level (such as gears) or a system level (such as CNC equipment, and conveyor belt). As the amount of generated failure data grows, it should help inform the overall production line operation for equipment health assessment.

Appendices

Appendix A   ISO Standards

Table A.1 contains a list of ISO standards from TC 108 and subcommittee 2 and 5.

Table A.1:   List of ISO condition monitoring standards from ISO TC 108 SC5. Italicized standards come from ISO TC 108 SC2.

ISO Standards Number	Title
ISO 13374-1: 2003	Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 1: General guidelines
ISO 13374-2: 2007	Condition monitoring and diagnostics of machines — Data processing, communication and presentation — Part 2: Data processing
ISO 22096: 2007	Condition monitoring and diagnostics of machines — Acoustic emission
ISO 18434-1: 2008	Condition monitoring and diagnostics of machines — Thermography — Part 1: General procedures
ISO 13374-3: 2012	Condition monitoring and diagnostics of machines — Data processing, communication and presentation — Part 3: Communication
ISO 13372: 2012	Condition monitoring and diagnostics of machines — Vocabulary
ISO 13379-1: 2012	Condition monitoring and diagnostics of machines — Data interpretation and diagnostics techniques — Part 1: General guidelines
ISO 18436-3: 2012	Condition monitoring and diagnostics of machines — Requirements for qualification and assessment of personnel — Part 3: Requirements for training bodies and the training process
ISO 18436-5: 2012	Condition monitoring and diagnostics of machines — Requirements for qualification and assessment of personnel — Part 5: Lubricant laboratory technician/analyst
ISO 18436-8: 2013	Condition monitoring and diagnostics of machines — Requirements for qualification and assessment of personnel — Part 8: Ultrasound
ISO 20958: 2013	Condition monitoring and diagnostics of machine systems — Electrical signature analysis of three-phase induction motors
ISO 18436-2: 2014	Condition monitoring and diagnostics of machines — Requirements for qualification and assessment of personnel — Part 2: Vibration condition monitoring and diagnostics

ISO 18436-4:2014	Condition monitoring and diagnostics of machines — Requirements for qualification and assessment of personnel — Part 4: Field lubricant analysis
ISO 18436-7: 2014	Condition monitoring and diagnostics of machines — Requirements for qualification and assessment of personnel — Part 7: Thermography
ISO 13379-2: 2015	Condition monitoring and diagnostics of machines — Data interpretation and diagnostics techniques — Part 2: Data-driven applications
ISO 13374-4: 2015	Condition monitoring and diagnostics of machine systems — Data processing, communication and presentation — Part 4: Presentation
ISO 13381-1: 2015	Condition monitoring and diagnostics of machines — Prognostics — Part 1: General guidelines
ISO 18129: 2015	Condition monitoring and diagnostics of machines — Approaches for performance diagnosis
ISO 16079-1: 2017	Condition monitoring and diagnostics of wind turbines — Part 1: General guidelines
ISO 17359: 2018	Condition monitoring and diagnostics of machines — General guidelines
ISO 18095: 2018	Condition monitoring and diagnostics of power transformers
ISO 29821: 2018	Condition monitoring and diagnostics of machines — Ultrasound — General guidelines, procedures and validation
ISO 14830-1: 2019	Condition monitoring and diagnostics of machine systems — Tribology-based monitoring and diagnostics — Part 1: General requirements and guidelines
ISO 18434-2: 2019	Condition monitoring and diagnostics of machine systems — Thermography — Part 2: Image interpretation and diagnostics
ISO 16079-2: 2020	Condition monitoring and diagnostics of wind turbines — Part 2: Monitoring the drivetrain
ISO 19283: 2020	Condition monitoring and diagnostics of machines — Hydroelectric generating units

ISO 18436-1: 2021	Condition monitoring and diagnostics of machine systems — Requirements for certification of personnel — Part 1: Sector specific requirements for certification bodies and the certification process
ISO 18436-6: 2021	Condition monitoring and diagnostics of machines — Requirements for certification of personnel — Part 6: Acoustic emission
ISO 13373-1: 2002	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 1: General procedures
ISO 13373-3: 2015	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 3: Guidelines for vibration diagnosis
ISO 13373-2: 2016	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 2: Processing, analysis and presentation of vibration data
ISO 13373-7: 2017	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 7: Diagnostic techniques for machine sets in hydraulic power generating and pump-storage plants
ISO 13373-9: 2017	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 9: Diagnostic techniques for electric motors
ISO 13373-5: 2020	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 5: Diagnostic techniques for fans and blowers
ISO 13373-4: 2021	Condition monitoring and diagnostics of machines — Vibration condition monitoring — Part 4: Diagnostic techniques for gas and steam turbines with fluid-film bearings

Appendix B Condition Monitoring Literature Review from Section 2

This section comprises the methodology for the literature search conducted in Section 2. The search begins by investigating the different data acquisition methods used in bearing condition monitoring. These terms are organized as follows:

1. ("Ball Bearing" and "Vibration")

2. ("Rolling Element Bearing" and "Vibration")

3. ("Ball Bearing" and "Acoustic Emission")

4. ("Rolling Element Bearing" and "Acoustic Emission")

5. ("Ball Bearing" and "Current")

6. ("Rolling Element Bearing" and "Current")

7. ("Ball Bearing" and "Sound")

8. ("Rolling Element Bearing" and "Sound")

9. ("Ball Bearing" and "Temperature")

10. ("Rolling Element Bearing" and "Temperature")

11. ("Ball Bearing" and "Ultrasonic")

12. ("Rolling Element Bearing" and "Ultrasonic")

These terms are comprised in the initial Engineering Village Search labeled in the list at the end of the section. Four filters were then added to the search:

1. English-only articles

2. Journal articles

3. Articles from the Compendex database

4. Articles from the time period of 2018 - 2023

After these initial search terms, a set of controlled vocabulary was selected to focus on applications related to bearings. The set of controlled vocab had similar terms across all three searches related to bearings. The final filtering focused on selecting articles only from the top 10 listed journals and the top five publishers. This helped identify the first search input. Subsequent search inputs were considered due to the lack of articles discovered in the initial search for ultrasonic sensing and the low number of articles for current sensing. A similar process was conducted for these data acquisition methods respectively. The search for all data acquisition articles yielded 170 articles. The search for the ultrasonic data acquisition articles yielded three articles and for current data acquisition yielded 30 articles. A total of 159 articles met the search criteria for inclusion in the review and are listed as such in Table 2.3.

The final completed terms for the three searches are listed below. The first search term comprised all of the condition monitoring technologies. The second considered only the ultrasonic data acquisition methods. The third search considered only the current sensing data acquisition methods.

1. **Initial Engineering Village Search:** ((((((((((((((((((((((((("Ball Bearing" and "Vibration")) WN ALL) OR ((("Rolling Element Bearing" and "Vibration")) WN ALL) OR ((("Ball Bearing" and "Acoustic Emission")) WN ALL)) OR ((("Rolling Element Bearing" and "Acoustic Emission")) WN ALL) OR ((("Ball Bearing" and "Current")) WN ALL)) OR ((("Rolling Element Bearing" and "Current")) WN ALL) OR ((("Ball Bearing" and "Sound")) WN ALL)) OR ((("Rolling Element Bearing" and "Sound")) WN ALL) OR ((("Rolling Element Bearing" and "Temperature")) WN ALL) OR ((("Ball Bearing" and "Temperature")) WN ALL) OR ((("Rolling Element Bearing" and "Ultrasonic")) WN ALL) OR ((("Ball Bearing" and "Ultrasonic")) WN ALL))) AND (ja WN DT)) AND (english WN LA)) AND ((2023 OR 2022 OR 2021 OR 2020 OR 2019 OR 2018) WN YR))))) AND ((cpx or c84 OR ins or ibf) WN DB))))) AND ((cpx or c84) WN DB)) AND (roller bearings OR ball bearings OR rolling bearings OR machine bearings OR rotors OR machine tool spindles OR wind turbines OR induction motors OR aircraft engines OR magnetic bearings OR rotors (mechanical) OR engines OR stators) WN CV))) AND ((english WN LA) AND ((mechanical systems and signal processing OR measurement: journal of the international measurement confederation OR sensors OR shock and vibration OR journal of sound and vibration OR measurement science and technology OR ieee access OR journal of vibroengineering OR proceedings of the institution of mechanical

393

engineers, part k: journal of multi-body dynamics OR sensors (switzerland)) WN ST))) AND ((academic press OR elsevier b.v. OR hindawi limited OR institute of electrical and electronics engineers inc. OR mdpi ag) WN PN))

2. **Individual Ultrasonic Engineering Village Search:** (((((((((((("Rolling Element Bearing" and "Ultrasonic") WN ALL) OR (("Ball Bearing" and "Ultrasonic") WN ALL)) OR (("Rolling Element Bearing" and "Ultrasound") WN ALL)) OR (("Ball Bearing" and "Ultrasound") WN ALL))) AND ((ball bearings OR roller bearings OR rolling bearings OR gears OR bearings (machine parts) OR stators OR bearings OR rotors OR stepping motors OR aircraft engines OR engines OR gas turbines OR rotating machinery OR turbomachine blades) WN CV)) AND (english WN LA)) AND (((cpx or c84) WN DB) AND (ja WN DT))) AND ((tribology transactions OR journal of the acoustical society of america OR journal of tribology OR mechanism and machine theory OR sensors and actuators, a: physical OR tribology international OR acs applied materials and interfaces OR applied energy OR applied mathematical modelling OR carbon) WN ST)) AND ((elsevier ltd OR taylor and francis inc. OR acoustical society of america OR american chemical society OR american society of mechanical engineers) WN PN)) AND ((2023 OR 2021 OR 2019) WN YR))

3. **Individual Motor Current Engineering Village Search:** ((((((((("Rolling Element Bearing" and "Current") WN ALL) OR (("Ball Bearing" and "Current") WN ALL))))) AND (((cpx or c84) WN DB) AND ((roller bearings OR ball bearings OR rolling bearings OR machine bearings OR electric motors OR rotating machinery OR bearings OR gas turbines OR rotors OR aircraft engines OR bearings (machine parts) OR magnetic bearings OR gears OR squirrel cage motors OR fighter aircraft OR machine tools OR engines OR electric machinery OR electric machines OR brushless dc motors OR electric drives OR pumps OR rotors (mechanical)) WN CV) AND ((2023 OR 2022 OR 2021 OR 2020 OR 2019 OR 2018) WN YR))) AND ((ja WN DT) AND (english WN LA) AND ((journal of tribology OR mechanical systems and signal processing OR measurement: journal of the international measurement confederation OR proceedings of the institution of mechanical engineers, part k: journal of multi-body dynamics OR tribology international OR journal of engineering for gas turbines and power OR lecture notes in mechanical engineering OR lubrication science OR mechanisms and machine science OR nano energy) WN ST) AND ((institute of electrical and electronics

engineers inc. OR american society of mechanical engineers (asme) OR elsevier ltd OR sage publications ltd OR elsevier b.v.) WN PN)))

Appendix C   Condition Monitoring Assessment

Table C.1 below contains a link for the bearing dataset locations (current as of Feburary 2023).

Table C.1: Bearing Dataset Links

Bearing dataset	Dataset link
IMS (NASA)	https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository
FEMTO (IEEE 2012 PHM)	https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository
CWRU	https://engineering.case.edu/bearingdatacenter
MFPT	https://www.mfpt.org/fault-data-sets/
UO	https://data.mendeley.com/datasets/v43hmbwxpm/2
PU	https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download
XJTU-SY	https://github.com/WangBiaoXJTU/xjtu-sy-bearing-datasets
KAU	https://www.KAU.com/bearing

Table C.2 contains the dataset characterisitics from Pipino *et al.* [2], Lessmeier *et al.* [413], and Hagmeyer *et al.* [416].

Table C.2: Data Assessment Criteria from Pipino *et al.*, Lessmeier *et al.*, and Hagmeyer *et al.*

Ref.	Pipino *et al.* [2]	Lessmeier *et al.* [413]	Hagmeyer *et al.* [416]
Criteria	Accessibility, Appropriate Amount of Data, Believability, Completeness, Concise Representation, Consistent Representation, Ease of Manipulation, Free-of-Error, Interpretability, Objectivity, Relevancy, Reputation, Security, Timeliness, Understandability, Value-Added	General Info, Manufacturer Information, Application Information, Damage Information	History Provided by Data, Data Acquisition, Data on degradation, System Information, Quality and Quantity of data, Data Source

Appendix D   Additional Chapter 5 Figures

D.1   Silhouette Score

Equation D.1 provides the method for calculating the silhouette score averaged over the number of samples. s is the silhouette score to describe how well a model fits defined clusters. a is the mean distance between a sample and other data points. b is the mean distance between a sample and the next nearest cluster. Figure D.1 provides a silhouette score example that illustrates Equation D.1.

$$s = \frac{(b - a)}{max(a, b)} \tag{D.1}$$
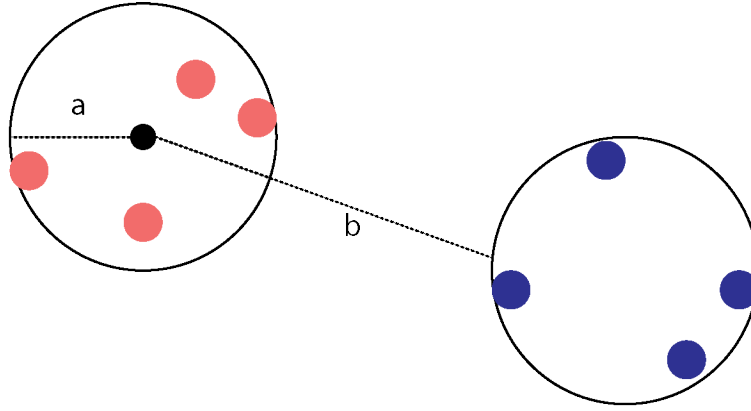
Figure D.1: Silhouette score example with supporting distances

D.2   Damage Methods

Figure D.2 shows the tool representation from Solidworks 2020. The tool cut away represents the area in contact with the bearing surface during the defect generation process.

D.3   Data Generation Figures

Figure D.3 shows the Kurtosis and skewness response for all the cases when all bearing data are present. For the 6205 bearing data, the inconsistencies in the skewness and kurtosis were noted in the initial characterization of the bearing baseline data (Section 4.3.1). However, based on the representation in the IR defect data, these inconsistencies may be anomalous based on poor operating
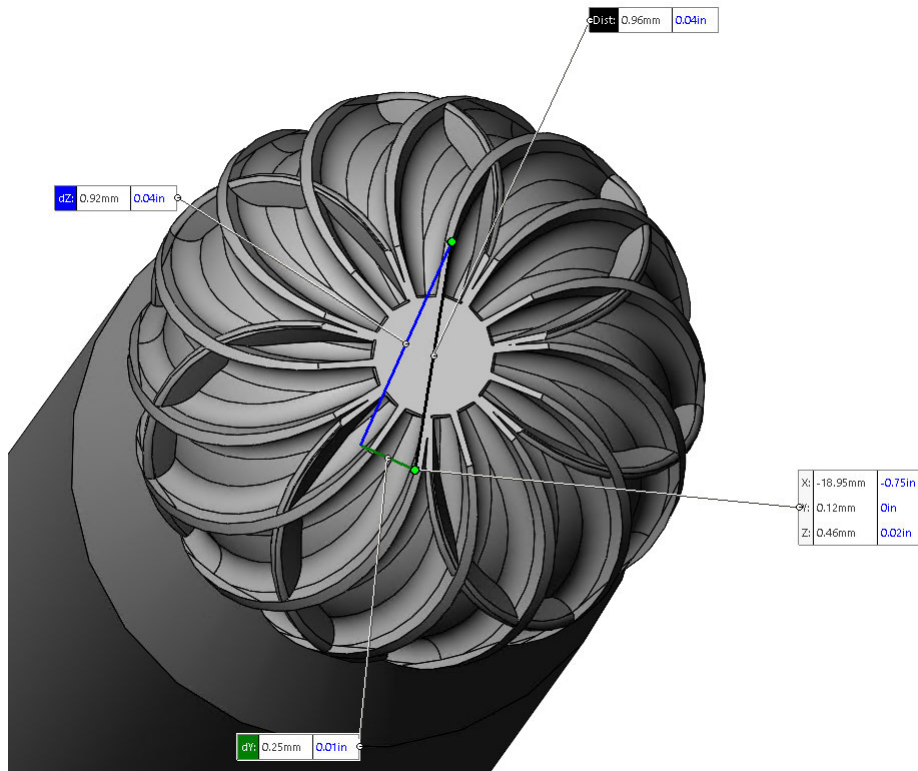
Figure D.2: Tool bit representation from Solidworks 2020

practices or the inability to isolate the bearing signal. Further examination of the frequency data could support the conclusions by possibly identifying which frequency ranges are causing the issues in the bearing data separation between features.

Figure D.4 and Figure D.5 show the IMF and HHT representations for the 6206 for another case of the baseline data. While the baseline is still greater than the inner race defect in terms of amplitude in the low-frequency signals. However, there is a closer representation between the baseline and inner race cases.
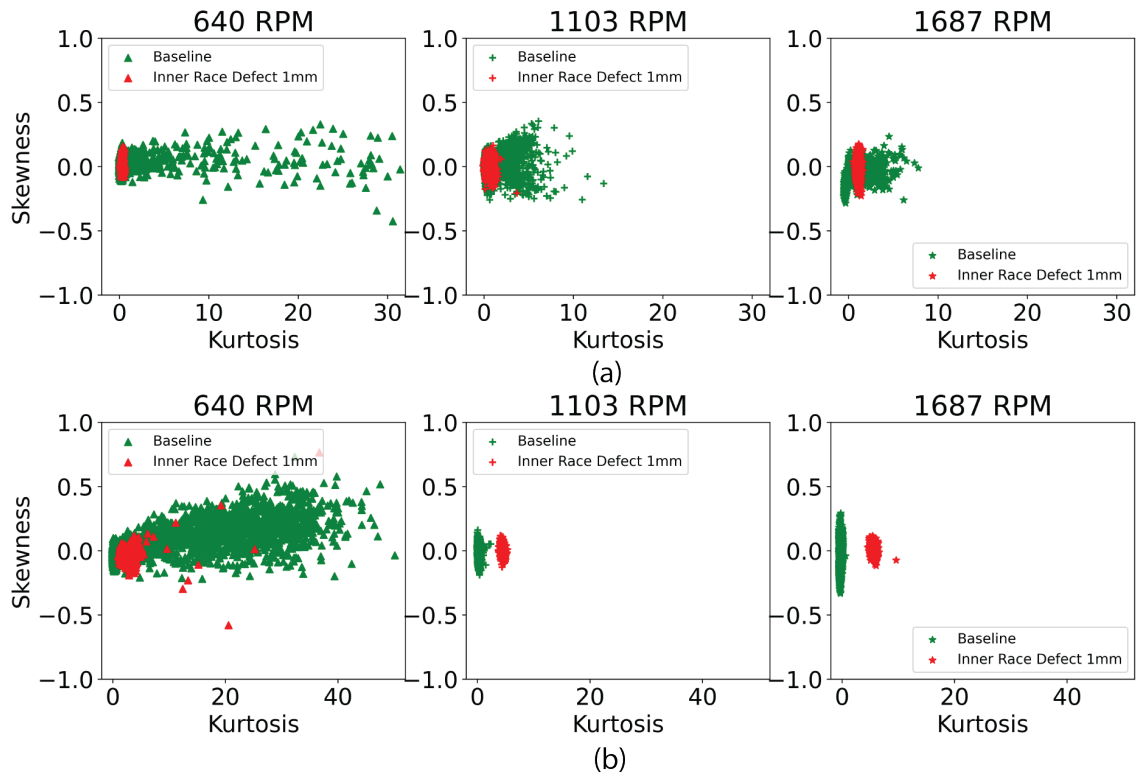
Figure D.3: Kurtosis and Skewness comparison between baseline and 1mm IR defect for the (a) 6205 and (b) 6207.
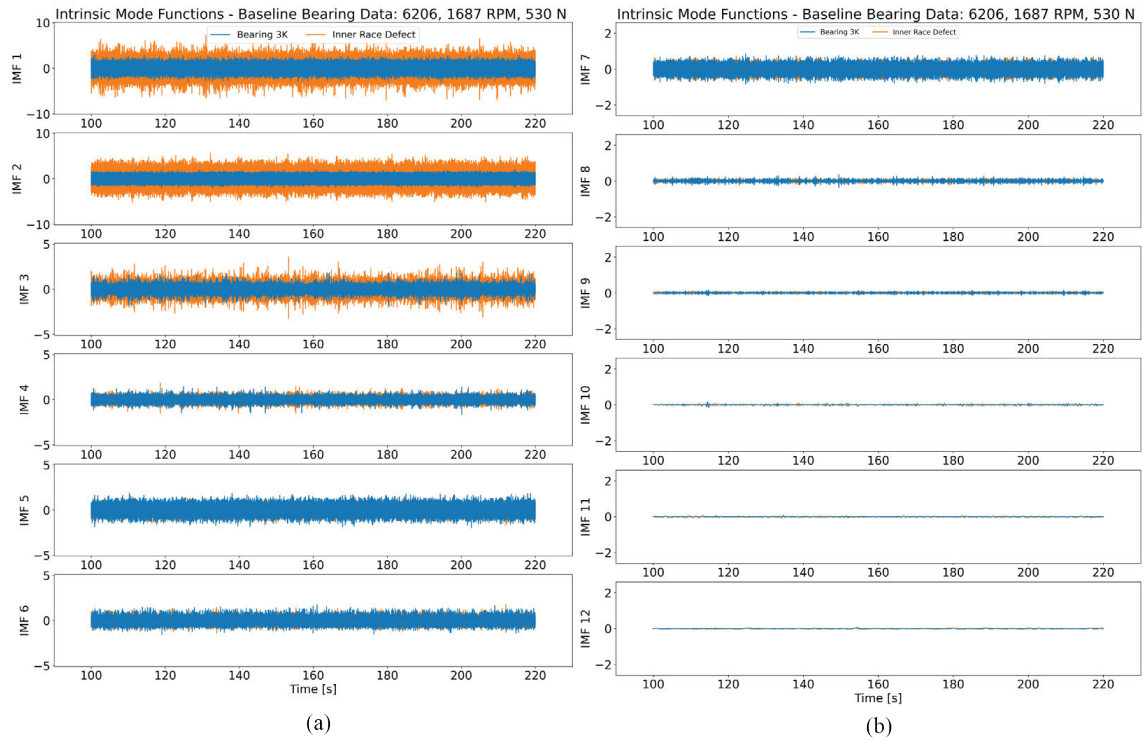
Figure D.4: 6206 Bearing 3K IMFs compared to inner race defect bearing at 1687 RPM
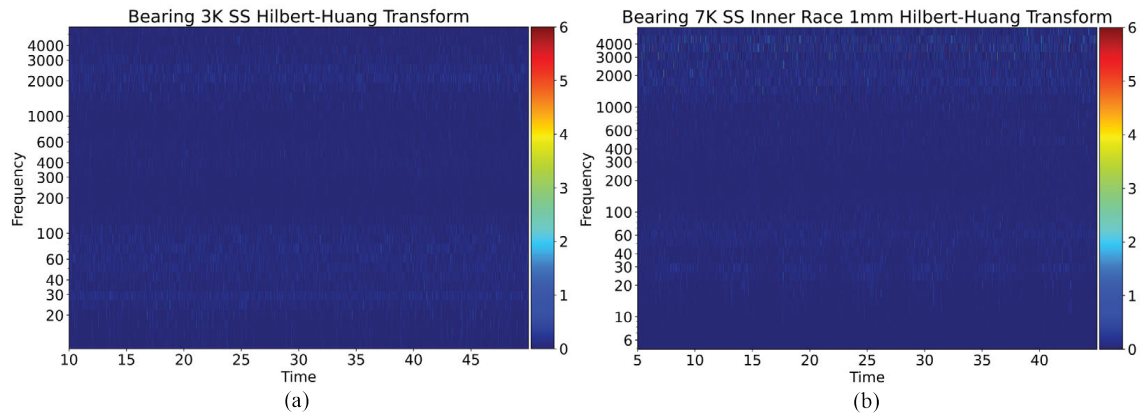
Figure D.5: 6206 Bearing 3K HHT compared to inner race defect bearing at 1687 RPM

Appendix E Additional Physics Model Info Information

E.1   Bearing Physical Parameters

Table E.1 contains bearing dimensionless parameters derived by Harris *et al.* [498]. These parameters are substituted into Equation 5.17 to calculate the actual dimensions to describe the bearing contact stresses and deformations. These parameters are derived based on the method used by Brewe *et al.* [515].

Table E.1: Bearing Dimensionless Parameters from Harris *et al.*

$F(\rho)$	a*	b*	$\delta*$
0	1	1	1
0.1075	1.0760	0.9318	0.9974
0.3204	1.2623	0.8114	0.9761
0.4795	1.4556	0.7278	0.9429
0.5916	1.6440	0.6687	0.9077
0.6716	1.8258	0.6245	0.8733
0.7332	2.011	0.5881	0.8394
0.7984	2.265	0.5480	0.7961
0.8349	2.494	0.5186	0.7602
0.8736	2.800	0.4863	0.7169
0.9099	3.233	0.4499	0.6636
0.9365	3.738	0.4166	0.6112
0.9573	4.395	0.3830	0.5551
0.9729	5.267	0.3490	0.4960
0.9837	6.448	0.3150	0.4352
0.9909	8.062	0.2814	0.3745
0.9951	10.222	0.2497	0.3176
0.9973	12.789	0.2232	0.2705
0.9981	14.839	0.2072	0.2472
0.9989	17.974	0.1882	0.2106
0.9994	23.55	0.1644	0.1716
0.9998	37.38	0.1305	0.1199
1	inf	0	0

Figure E.1 provides the IMF energy levels for the different IMFs for the 6206 case. The different bearings have different shapes that also appeared to be retained during the spectral augmentation process.

(a)

(b)

(c)

(d)

(e)

(f)
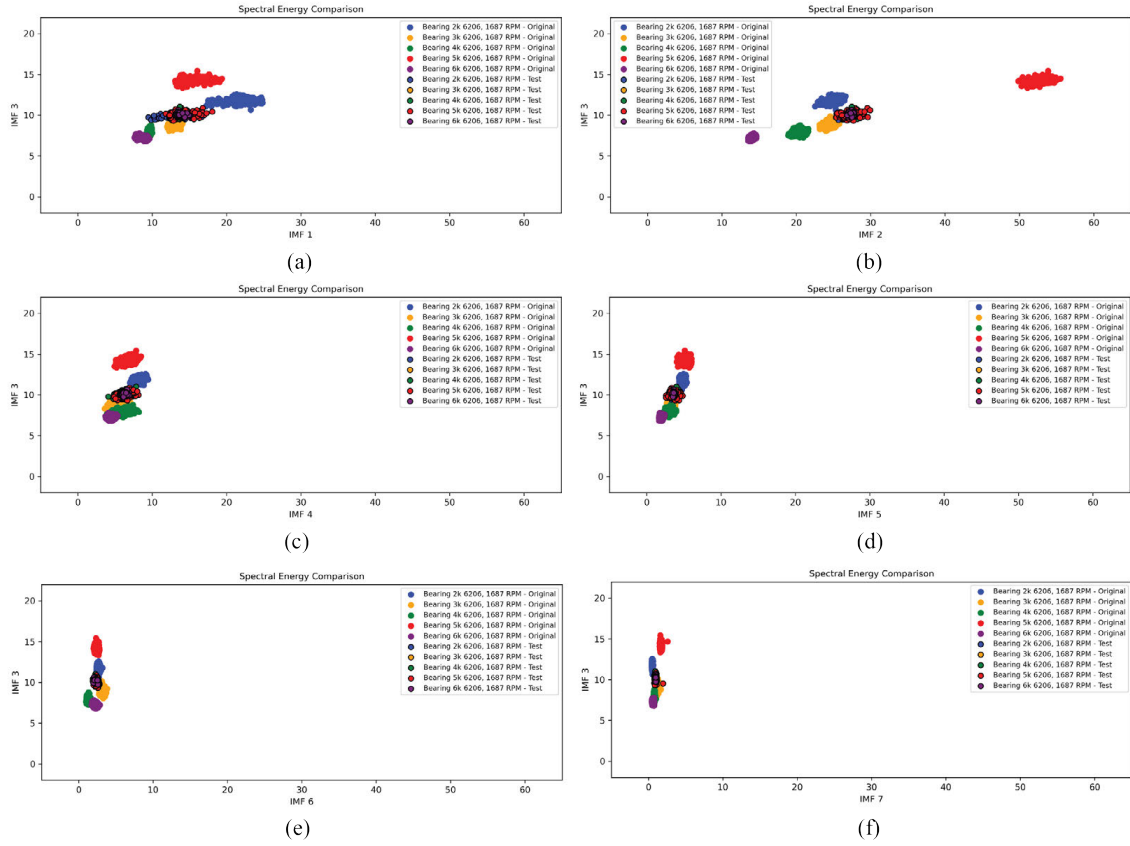
Figure E.1: 6206 baseline IMF energy levels for the spectral augmentation. (a) IMF 1 vs. IMF 3, (b) IMF 2 vs. IMF 3, (c) IMF 4 vs. IMF 3, (d) IMF 5 vs. IMF 3, (e) IMF 6 vs. IMF 3, (f) IMF 7 vs. IMF 3

Appendix F Additional Chapter 6 Information

F.1 Activation Functions

As mentioned, there are four different activation functions: the hyperbolic tangent function, the sigmoid activation function, the rectified linear unit activation function (relu), and the leaky rectified linear unit activation. A fifth activation function is considered as the identity. This function is used instead of the leaky relu due to the scikit-learn implementation.

$$f(x) = \frac{exp2x - 1}{exp2x + 1} \tag{F.1}$$

$$f(x) = \frac{1}{1 + exp-x} \tag{F.2}$$

$$f(x) = max(0, x) \tag{F.3}$$

$$f(x) = max(0.1x, x) \tag{F.4}$$

$$f(x) = x \tag{F.5}$$

Bibliography

[1] Thomas DS, Weiss BA, Thomas DS. Economics of Manufacturing Machinery Maintenance A Survey and Analysis of U . S . Costs and Benefits. NIST; 2020.

[2] Pipino LL, Lee YW, Wang RY, Yang RY. Data Quality Assessment. Communications of the ACM. 2002;45(4):211-8.

[3] Wescoat E, Krugh M, Jansari V, Mears L. Redefining the Digital Triplet for Surrogate System Integration. Manufacturing Letters. 2023;In Press. Available from: `https://www.sciencedirect.com/science/article/pii/S2213846323000159`.

[4] Pahl G, Beitz W, Feldhusen J, Grote KH. Engineering Design: A Systematic Approach. 3rd ed. Berlin Heidelberg: Springer; 2007.

[5] ISO. ISO 17359: Condition Monitoring and diagnostics of machines - general guidelines. Geneva: International Organization of Standardization; 2018.

[6] ISO. ISO 15243:2017 - Rolling Bearings - Damage and Failures. ISO; 2017.

[7] SKF. Bearing damage and failure analysis; 2017. Available from: `https://www.skf.com/binaries/pub12/Images/0901d1968064c148-Bearing-failures---14219_2-EN_tcm_12-297619.pdf`.

[8] Mohajan H. The First Industrial Revolution: Creation of a New Global Human Era. Journal of Social Sciences and Humanities. 2019;5(4):377-87. Available from: `https://mpra.ub.uni-muenchen.de/96644/`.

[9] Mohajan H. The Second Industrial Revolution has Brought Modern Social and Economic Developments. Journal of Social Sciences and Humanities. 2020;6(1):1-14.

[10] Ghislieri C, Molino M, Cortese CG. Work and organizational psychology looks at the Fourth Industrial Revolution: How to support workers and organizations? Frontiers in Psychology. 2018;28(NOV):1-6.

[11] Xu M, David JM, Kim SH. The fourth industrial revolution: Opportunities and challenges. International Journal of Financial Research. 2018;9(2):90-5.

[12] Sarfraz Z, Sarfraz A, Iftikar HM, Akhund R. Is COVID-19 pushing us to the Fifth Industrial Revolution (Society 5.0)? Pak J Med Sci. 2021;37(2):591-4.

[13] Sagiroglu S, Sinanc D. Big Data : A Review. In: 2013 International Conference on Collaboration Technologies and Systems (CTS). San Diego: IEEE; 2013. p. 42-7.

[14] Lee J, Bagheri B, Kao HA. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters. 2015;3:18-23. Available from: `http://dx.doi.org/10.1016/j.mfglet.2014.12.001`.

[15] Haag S, Anderl R. Digital twin – Proof of concept. Manufacturing Letters. 2018;15:64-6. Available from: https://doi.org/10.1016/j.mfglet.2018.02.006.

[16] Liu Y, Zhang L, Yang Y, Zhou L, Ren L, Wang F, et al. A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin. IEEE Access. 2019;7:49088-101.

[17] Fan C, Zhang C, Yahja A, Mostafavi A. Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. International Journal of Information Management. 2021;56(March 2019):102049. Available from: https://doi.org/10.1016/j.ijinfomgt.2019.102049.

[18] Almeaibed S, Al-Rubaye S, Tsourdos A, Avdelidis NP. Digital Twin Analysis to Promote Safety and Security in Autonomous Vehicles. IEEE Communications Standards Magazine. 2021;5(1):40-6.

[19] Trauer J, Schweigert-Recksiek S, Engel C, Spreitzer K, Zimmermann M. What is a Digital Twin? Definitions and Insights from an Industrial Case Study in Technical Product Development. In: Proceedings of the Design Society: DESIGN Conference 2020 May; 2020. p. 757-66.

[20] Monostori L, Kadar B, Bauernhansl T, Kondoh S, Kumara S, Reinhart G, et al. Cyber-physical systems in manufacturing. CIRP Annals - Manufacturing Technology. 2016;65(2016):621-41.

[21] Palensky P, Member S, Widl E, Elsheikh A. Simulating Cyber-Physical Energy Systems: Challenges, Tools and Methods. IEEE Transactions on Systems, Man, and Cybernetrics: Systems. 2014;44(3):318-26.

[22] Vogl GW, Weiss BA, Helu M. A review of diagnostic and prognostic capabilities and best practices for manufacturing. Journal of Intelligent Manufacturing. 2019;30(1):79-95.

[23] Leukel J, González J, Riekert M. Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. Journal of Manufacturing Systems. 2021 oct;61:87-96.

[24] Vieira GE, Herrmann JW, Lin E. Rescheduling manufacturing systems: A framework of strategies, policies, and methods. Journal of Scheduling. 2003;6(1):39-62.

[25] Pujawan IN, Bah AU. Supply chains under COVID-19 disruptions: literature review and research agenda. Supply Chain Forum. 2022;23(1):81-95. Available from: https://doi.org/10.1080/16258312.2021.1932568.

[26] Moosavi J, Fathollahi-Fard AM, Dulebenets MA. Supply chain disruption during the COVID-19 pandemic: Recognizing potential disruption management strategies. International Journal of Disaster Risk Reduction. 2022;75(April):102983. Available from: https://doi.org/10.1016/j.ijdrr.2022.102983.

[27] Li HY, Xu W, Cui Y, Wang Z, Xiao M, Sun ZX. Preventive Maintenance Decision Model of Urban Transportation System Equipment Based on Multi-Control Units. IEEE Access. 2020;8:15851-69.

[28] Maktoubian J, Ansari K. An IoT architecture for preventive maintenance of medical devices in healthcare organizations. Health and Technology. 2019;9(3):233-43.

[29] Ahmad T, Zhang D, Huang C, Zhang H, Dai N, Song Y, et al. Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. Journal of Cleaner Production. 2021;289:125834. Available from: https://doi.org/10.1016/j.jclepro.2021.125834.

[30] Heneghan C, Thompson M, Billingsley M, Cohen D. Medical-device recalls in the UK and the device-regulation process: retrospective review of safety notices and alerts. BMJ Open. 2011;1(1):e000155-5.

[31] Sánchez-Barroso G, Sanz-Calcedo JG. Application of predictive maintenance in hospital heating, ventilation and air conditioning facilities. Emerging Science Journal. 2019;3(5):337-43.

[32] Gerdes M, Scholz D, Galar D. Effects of condition-based maintenance on costs caused by unscheduled maintenance of aircraft. Journal of Quality in Maintenance Engineering. 2016;22(4):394-417.

[33] FAA. FAA, editor. Air Traffic By the Numbers. FAA; 2023. Available from: `https://www.faa.gov/air_traffic/by_the_numbers`.

[34] Chong AKW, Mohammed AH, Abdullah MN, Rahman MSA. Maintenance prioritization – a review on factors and methods. Journal of Facilities Management. 2019;17(1):18-39.

[35] Muller MR, Kasten D. Industry Productivity Training Manual. New Brunswick: Rutgers, The State University of New Jersey Center for Advanced Energy Systems; 2006.

[36] Ravande S. Forbes, editor. Unplanned Downtime Costs More Than You Think. Forbes; 2022. Available from: `https://www.forbes.com/sites/forbestechcouncil/2022/02/22/unplanned-downtime-costs-more-than-you-think/?sh=14f751de36f7`.

[37] Rao P, Sheaffer P, Chen Y, Goldberg M, Jones B, Cropp J, et al. U.S. Industrial and Commercial Motor System Market Assessment Report Volume 1: Characteristics of the Installed Base. Berkeley: Energy Technologies Area, Lawrence Berkeley National Laboratory; 2021. January. Available from: `https://escholarship.org/uc/item/42f631k3`.

[38] O'Donnell P, Heising C, Singh C, Wells SJ. Report of Large Motor Reliability Survey of Industrial and Commercial Installations, Part I. IEEE Transactions on Industry Applications. 1985;IA-23(4):153-8.

[39] O'Donnell P, Heising C, Singh C, Wells SJ. Report of Large Motor Reliability Survey of Industrial and Commercial Installations, Part II. IEEE Transactions on Industry Applications. 1985;IA-23(4):153-8.

[40] O'Donnell P, Heising C, Singh C, Wells SJ. Report of Large Motor Reliability Survey of Industrial and Commercial Installations: Part 3. IEEE Transactions on Industry Applications. 1987;IA-23(1):153-8.

[41] Liu W, Zhang S, Lin J, Xia Y, Wang J, Sun Y. Advancements in accuracy decline mechanisms and accuracy retention approaches of CNC machine tools: a review. International Journal of Advanced Manufacturing Technology. 2022;121(11-12):7087-115. Available from: `https://doi.org/10.1007/s00170-022-09720-0`.

[42] Guo Y, Sheng S, Phillips C, Keller J, Veers P, Williams L. A methodology for reliability assessment and prognosis of bearing axial cracking in wind turbine gearboxes. Renewable and Sustainable Energy Reviews. 2020;127(April):109888. Available from: `https://doi.org/10.1016/j.rser.2020.109888`.

[43] Kim HE, Tan ACC, Mathew J, Choi BK. Bearing fault prognosis based on health state probability estimation. Expert Systems with Applications. 2012;39(5):5200-13. Available from: `http://dx.doi.org/10.1016/j.eswa.2011.11.019`.

[44] Prinzinger JM, Rickmeyer TC. Summary of U.S. Army Seeded Fault Tests for Helicopter Bearings. Redstone Arsenal: US Department of Transportation Federal Aviation Administration; 2016.

[45] Eti MC, Ogaji SOT, Probert SD. Reducing the cost of preventive maintenance (PM) through adopting a proactive reliability-focused culture. Applied Energy. 2006;83(11):1235-48.

[46] Walford CA. Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs. Albuquerque: Sandia National Laboratories; 2006. Available from: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online.

[47] Sheng S. Report on Wind Turbine Subsystem Reliability A Survey of Various Databases. Golden: National Renewable Neergy Laboratory; 2013. Available from: https://www.nrel.gov/docs/fy13osti/59111.pdf.

[48] EPA. Chevron Phillips Chemical Company Clean LP Air Act Settlement; 2022. Available from: https://www.epa.gov/enforcement/chevron-phillips-chemical-company-clean-lp-air-act-settlement.

[49] Debusmann Jr B. Ohio train derailment: Norfolk Southern ordered to pay for cleanup costs. Washington, D.C.; 2023. Available from: https://www.bbc.com/news/world-us-canada-64722765.

[50] Cohen L. Ohio Gov. says Norfolk Southern "needs to pay" for health care costs related to East Palestine trail derailment. CBS; 2023. Available from: https://www.cbsnews.com/news/ohio-train-derailment-mike-dewine-epa-norfolk-southern-pay-health-care-costs/.

[51] Kalpajian S, Schmid S. Manufacturing Processes for Engineering Materials. 5th ed. Pearson Education; 2008.

[52] Dunning KK, Davis KG, Cook C, Kotowski SE, Hamrick C, Jewell G, et al. Costs by industry and diagnosis among musculoskeletal claims in a state workers compensation system: 1999-2004. American Journal of Industrial Medicine. 2010;53(3):276-84.

[53] Poor P, Basl J, Zenisek D. Predictive Maintenance 4.0 as next evolution step in industrial maintenance development. In: 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE). University of Kelaniya; 2019. p. 245-53.

[54] Cooke FL. Plant maintenance strategy: Evidence from four British manufacturing firms. Journal of Quality in Maintenance Engineering. 2003;9(3):239-49.

[55] Swanson L. Linking maintenance strategies to performance. International Journal of Production Economics. 2001;70(3):237-44.

[56] Wu S, Clements-Croome D. Preventive maintenance models with random maintenance quality. Reliability Engineering and System Safety. 2005;90(1):99-105.

[57] Canfield RV. Cost Optimization of Periodic Preventive Maintenance. IEEE Transactions on Reliability. 1986;35(1):78-81.

[58] Wu S, Zuo MJ. Linear and nonlinear preventive maintenance models. IEEE Transactions on Reliability. 2010;59(1):242-9.

[59] Xiaofei L, Min L. Hazard rate function in dynamic environment. Reliability Engineering and System Safety. 2014;130:50-60. Available from: http://dx.doi.org/10.1016/j.ress.2014.04.020.

[60] Yang L, Zhao Y, Peng R, Ma X. Hybrid preventive maintenance of competing failures under random environment. Reliability Engineering and System Safety. 2018;174(September 2017):130-40. Available from: https://doi.org/10.1016/j.ress.2018.02.017.

[61] Kothamasu R, Huang SH, Verduin WH. System health monitoring and prognostics - A review of current paradigms and practices. International Journal of Advanced Manufacturing Technology. 2006;28(9):1012-24.

[62] Niu G, Yang BS, Pecht M. Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. Reliability Engineering and System Safety. 2010;95(7):786-96. Available from: http://dx.doi.org/10.1016/j.ress.2010.02.016.

[63] Zonta T, da Costa CA, da Rosa Righi R, de Lima MJ, da Trindade ES, Li GP. Predictive maintenance in the Industry 4.0: A systematic literature review. Computers and Industrial Engineering. 2020 dec;150.

[64] Zhang W, Yang D, Wang H. Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. IEEE Systems Journal. 2019;13(3):2213-27.

[65] Arena F, Collotta M, Luca L, Ruggieri M, Termine FG. Predictive Maintenance in the Automotive Sector: A Literature Review. Mathematical and Computational Applications. 2021;27(1):2.

[66] Aivaliotis P, Arkouli Z, Georgoulias K, Makris S. Degradation curves integration in physics-based models: Towards the predictive maintenance of industrial robots. Robotics and Computer-Integrated Manufacturing. 2021;71:102177. Available from: https://doi.org/10.1016/j.rcim.2021.102177.

[67] Cao Q, Zanni-Merk C, Samet A, Reich C, de Beuvron FdB, Beckmann A, et al. KSPMI: A Knowledge-based System for Predictive Maintenance in Industry 4.0. Robotics and Computer-Integrated Manufacturing. 2022;74(November 2021).

[68] Adu-Amankwa K, Attia AKA, Janardhanan MN, Patel I. A predictive maintenance cost model for CNC SMEs in the era of industry 4.0. International Journal of Advanced Manufacturing Technology. 2019 oct;104(9-12):3567-87.

[69] de Pater I, Reijns A, Mitici M. Alarm-based predictive maintenance scheduling for aircraft engines with imperfect Remaining Useful Life prognostics. Reliability Engineering and System Safety. 2022;221(October 2021):108341. Available from: https://doi.org/10.1016/j.ress.2022.108341.

[70] Umeda S, Tamaki K, Sumiya M, Kamaji Y. Planned Maintenance Schedule Update Method for Predictive Maintenance of Smiconductor Plasma Etcher. IEEE Transactions on Semiconductor Manufacturing. 2021;34(3):296-300.

[71] Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing. 2006;20(7):1483-510.

[72] Inturi V, Sabareesh GR, Supradeepan K, Penumakala PK. Integrated condition monitoring scheme for bearing fault diagnosis of a wind turbine gearbox. JVC/Journal of Vibration and Control. 2019;25(12):1852-65.

[73] De Silva CW. Vibration: Fundamentals and Practice. 2nd ed. Taylor & Francis Group; 2006.

[74] Ogata K. System Dynamics. 3rd ed. Upper Saddle River: Pearson: Prentice Hall; 2004.

[75] Bogue R. Sensors for condition monitoring: A review of technologies and applications. Sensor Review. 2013;33(4):295-9.

[76] Haj Mohamad T, Nataraj C. Fault identification and severity analysis of rolling element bearings using phase space topology. JVC/Journal of Vibration and Control. 2021;27(3-4):295-310.

[77] Bernal E, Spiryagin M, Cole C. Onboard Condition Monitoring Sensors, Systems and Techniques for Freight Railway Vehicles: A Review. IEEE Sensors Journal. 2019;19(1):4-24.

[78] Huang Q, Tang B, Deng L. Development of high synchronous acquisition accuracy wireless sensor network for machine vibration monitoring. Measurement: Journal of the International Measurement Confederation. 2015;66:35-44. Available from: `http://dx.doi.org/10.1016/j.measurement.2015.01.021`.

[79] Parekh R. AC Induction Motor fundamentals. Microchip Technilogy Inc.; 2003.

[80] Avina-Corral V, Rangel-Magdaleno J, Morales-Perez C, Hernandez J. Bearing Fault Detection in Adjustable Speed Drive-Powered Induction Machine by Using Motor Current Signature Analysis and Goodness-of-Fit Tests. IEEE Transactions on Industrial Informatics. 2021;17(12):8265-74.

[81] Hall EH. On a New action of the Magnet on Electric currents. American Journal of Mathematics. 1879;2(3):287-92.

[82] Weiss R, Itzke A, Reitenspieb J, Hoffmann I, Weigel R. A Novel Closed Loop Current Sensor Based on a Circular Array of Magnetic Field Sensors. IEEE Sensors Journal. 2019;19(7):2517-24.

[83] Corne B, Vervisch B, Debruyne C, Knockaert J, Desmet J. Comparing MCSA with vibration analysis in order to detect bearing faults - A case study. Proceedings - 2015 IEEE International Electric Machines and Drives Conference, IEMDC 2015. 2016:1366-72.

[84] Roberts TM, Talebzadeh M. Acoustic emission monitoring of fatigue crack propagation. Journal of Constructional Steel Research. 2003;59(6):695-712.

[85] Biswas AK, Kumar Datta A, Topdar P, Sengupta S. On Effective Placement of Acoustic Emission Sensor in Steel Framed Structure for Damage Detection. In: 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications, NCETSTEA 2020; 2020. p. 0-4.

[86] Poddar S, Tandon N. Study of oil starvation in journal bearing using acoustic emission and vibration measurement techniques. Journal of Tribology. 2020;142(12):1-11.

[87] Poddar S, Tandon N. Detection of particle contamination in journal bearing using acoustic emission and vibration monitoring techniques. Tribiology International. 2019;134(February):154-64. Available from: `https://doi.org/10.1016/j.triboint.2019.01.050`.

[88] Shiroishi J, Li Y, Liang S, Kurfess T, Danyluk S. Bearing Condition Diagnostics Via Vibration and Acoustic Emission Measurements. Mechanical Systems and Signal Processing. 1997;11(5):693-705.

[89] Heng RBW, Nor MJM. Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition. Applied Acoustics. 1998;53(1-3):211-26.

[90] Mian T, Choudhary A, Fatima S. An efficient diagnosis approach for bearing faults using sound quality metrics. Applied Acoustics. 2022;195:108839. Available from: `https://doi.org/10.1016/j.apacoust.2022.108839`.

[91] Kim SJ, Kim S, Lee S, Youn BD, Kim T. Effective band-selection algorithm for rolling element bearing diagnosis using AE sensor data under noisy conditions. Structural and Multidisciplinary Optimization. 2022;65(9):1-23. Available from: `https://doi.org/10.1007/s00158-022-03360-4`.

[92] Chaturved GK, Thomas DW. Adaptive noise cancelling and condition monitoring. Journal of Sound and Vibration. 1981;76(3):391-405.

[93] Skoglind B, Roberts T, Karmakar S, Turner C, Mears L. Localized Acoustic-Event Measurement Probe: Connector Confirmation Utilizing Acoustic Signatures. In: Proceedings of the ASME 2020 15th International Manufacturing Science and Engineering Conference MSEC2020; 2020. p. 1-10.

[94] Detweiler WH. What's Normal: The Role of Temperature in Bearing Applications. Evolution: Technology Magazine from SKF. 1998 feb:1. Available from: `https://evolution.skf.com/us/whats-normalthe-role-of-temperature-in-bearing-applications-3/#:$\sim$:text=Normalmotorbearingoperatingtemperatures,estimatethecorrectbearingtemperature.`

[95] Draney RK. High temperature sensor for bearing health monitoring. In: IEEE Aerospace Conference Proceedings. Big Sky; 2008. p. 1-7.

[96] Barksdale H, Smith Q, Khan M. Condition Monitoring of Electrical Machines with Internet of Things. Conference Proceedings - IEEE SOUTHEASTCON. 2018;2018-April:3-6.

[97] Seebeck TJ. "Magnetische Polarisation der Metalle und Erze durch Temperatur-Differenz" [Magnetic polarization of metals and ores by temperature differences]. Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin. 1822:265-373.

[98] Danisman K, Dalkiran I, Celebi FV. Design of a high precision temperature measurement system based on artificial neural network for different thermocouple types. Measurement. 2006;39(8):695-700.

[99] Mayr J, Jedrzejewski J, Uhlmann E, Alkan Donmez M, Knapp W, Härtig F, et al. Thermal issues in machine tools. CIRP Annals - Manufacturing Technology. 2012;61(2):771-91.

[100] Albers T, Bonnett AH. Motor temperature considerations for pulp and paper mill applications. IEEE Transactions on Industry Applications. 2002;38(6):1701-13.

[101] Lu B, Li Y, Wu X, Yang Z. A review of recent advances in wind turbine condition monitoring and fault diagnosis. 2009 IEEE Power Electronics and Machines in Wind Applications, PEMWA 2009. 2009.

[102] Nembhard AD, Sinha JK, Pinkerton AJ, Elbhbah K. Combined vibration and thermal analysis for the condition monitoring of rotating machinery. Structural Health Monitoring. 2014;13(3):281-95.

[103] García Márquez FP, Tobias AM, Pinar Pérez JM, Papaelias M. Condition monitoring of wind turbines: Techniques and methods. Renewable Energy. 2012;46:169-78. Available from: `http://dx.doi.org/10.1016/j.renene.2012.03.003`.

[104] Leventhall G. What is infrasound? Progress in Biophysics and Molecular Biology. 2007;93(1-3):130-7.

[105] Kim EY, Tan ACC, Mathew J, Yang BS. Condition monitoring of low speed bearings: A comparative study of the ultrasound technique versus vibration measurements. Australian Journal of Mechanical Engineering. 2008;5(2):177-89.

[106] Drinkwater BW, Zhang J, Kirk KJ, Elgoyhen J, Dwyer-Joyce RS. Ultrasonic measurement of rolling bearing lubrication using piezoelectric thin films. Journal of Tribology. 2009;131(1):1-8.

[107] Li M, Jing M, Chen Z, Liu H. An improved ultrasonic method for lubricant-film thickness measurement in cylindrical roller bearings under light radial load. Tribology International. 2014;78:35-40. Available from: `http://dx.doi.org/10.1016/j.triboint.2014.04.023`.

[108] Zhang K, Meng Q, Chen W, Li J, Harper P. Ultrasonic measurement of oil film thickness between the roller and the inner raceway in a roller bearing. Industrial Lubrication and Tribology. 2015 sep;67(6):531-7.

[109] Scheeren B, Kaminski ML, Pahlavan L. Evaluation of Ultrasonic Stress Wave Transmission in Cylindrical Roller Bearings for Acoustic Emission Condition Monitoring. Sensors. 2022;22(4).

[110] Liu J, Xu Y, Pan G. A combined acoustic and dynamic model of a defective ball bearing. Journal of Sound and Vibration. 2021;501:116029. Available from: `https://doi.org/10.1016/j.jsv.2021.116029`.

[111] Luo Y, Chen C, Kang S, Zhang P. Fault Diagnosis of Rolling Element Bearing Using an Adaptive Multiscale Enhanced Combination Gradient Morphological Filter. Shock and Vibration. 2019;2019.

[112] Singh S, Howard CQ, Hansen CH, Köpke UG. Analytical validation of an explicit finite element model of a rolling element bearing with a localised line spall. Journal of Sound and Vibration. 2018;416:94-110. Available from: `http://dx.doi.org/10.1016/j.jsv.2017.09.007`.

[113] Jablonski A. Hybrid Model of Rolling-Element Bearing Vibration Signal. Energies. 2022;15(13).

[114] Mufazzal S, Muzakkir SM, Khanam S. Theoretical and experimental analyses of vibration impulses and their influence on accurate diagnosis of ball bearing with localized outer race defect. Journal of Sound and Vibration. 2021;513(August):116407. Available from: `https://doi.org/10.1016/j.jsv.2021.116407`.

[115] Motahari-Nezhad M, Jafari SM. ANFIS system for prognosis of dynamometer high-speed ball bearing based on frequency domain acoustic emission signals. Measurement: Journal of the International Measurement Confederation. 2020;166:108154. Available from: `https://doi.org/10.1016/j.measurement.2020.108154`.

[116] Soave E, D'Elia G, Cocconcelli M, Battarra M. Blind deconvolution criterion based on Fourier–Bessel series expansion for rolling element bearing diagnostics. Mechanical Systems and Signal Processing. 2022;169(October 2021):108588. Available from: `https://doi.org/10.1016/j.ymssp.2021.108588`.

[117] Luo M, Guo Y, Andre H, Wu X, Na J. Dynamic modeling and quantitative diagnosis for dual-impulse behavior of rolling element bearing with a spall on inner race. Mechanical Systems and Signal Processing. 2021;158:107711. Available from: `https://doi.org/10.1016/j.ymssp.2021.107711`.

[118] Liu J, Xu Z, Xu Y, Liang X, Pang R. An analytical method for dynamic analysis of a ball bearing with offset and bias local defects in the outer race. Journal of Sound and Vibration. 2019;461:114919. Available from: `https://doi.org/10.1016/j.jsv.2019.114919`.

[119] Xu L, Chatterton S, Pennacchi P. Rolling element bearing diagnosis based on singular value decomposition and composite squared envelope spectrum. Mechanical Systems and Signal Processing. 2021;148:107174. Available from: `https://doi.org/10.1016/j.ymssp.2020.107174`.

[120] Wan L, Chen Y, Li H, Li C. Rolling-element bearing fault diagnosis using improved lenet-5 network. Sensors (Switzerland). 2020;20(6):1-23.

[121] Xu Z, Tang G, He M. Peak-Based Mode Decomposition for Weak Fault Feature Enhancement and Detection of Rolling Element Bearing. Shock and Vibration. 2020;2020.

[122] Li Y, Zheng F, Xiong Q, Zhang W. A secondary selection-based orthogonal matching pursuit method for rolling element bearing diagnosis. Measurement: Journal of the International Measurement Confederation. 2021;176(February):109199. Available from: `https://doi.org/10.1016/j.measurement.2021.109199`.

[123] Liang L, Ding X, Liu F, Chen Y, Wen H. Feature extraction using sparse kernel non-negative matrix factorization for rolling element bearing diagnosis. Sensors. 2021;21(11).

[124] Wang G, Zhao B, Xiang L, Li W, Zhu C. Information interval spectrum: A novel methodology for rolling-element bearing diagnosis. Measurement: Journal of the International Measurement Confederation. 2021;183(June):109899. Available from: `https://doi.org/10.1016/j.measurement.2021.109899`.

[125] Huang W, Sun H, Luo J, Wang W. Periodic feature oriented adapted dictionary free OMP for rolling element bearing incipient fault diagnosis. Mechanical Systems and Signal Processing. 2019;126:137-60. Available from: `https://doi.org/10.1016/j.ymssp.2019.02.023`.

[126] Wu Y, Hu Z, An W, Li X, Wang X, Du P, et al. Automatic Diagnosis of Rolling Element Bearing under Different Conditions Based on RVMD and Envelope Order Capture. IEEE Access. 2019;7:91799-808.

[127] Jia M, Xu Y, Hong M, Hu X, Forte P. Multitask Convolutional Neural Network for Rolling Element Bearing Fault Identification. Shock and Vibration. 2020;2020.

[128] Li Q, Yan C, Wang W, Babiker A, Wu L. Health indicator construction based on MD-CUMSUM with multi-domain features selection for rolling element bearing fault diagnosis. IEEE Access. 2019;7:138528-40.

[129] Hou GQ, Lee CM. Estimation of the Defect Width on the Outer Race of a Rolling Element Bearing under Time-Varying Speed Conditions. Shock and Vibration. 2019;2019.

[130] Yan X, Liu Y, Jia M. A feature selection framework-based multiscale morphological analysis algorithm for fault diagnosis of rolling element bearing. IEEE Access. 2019;7:123436-52.

[131] Guo J, Zhen D, Li H, Shi Z, Gu F, Ball AD. Fault feature extraction for rolling element bearing diagnosis based on a multi-stage noise reduction method. Measurement: Journal of the International Measurement Confederation. 2019;139:226-35. Available from: `https://doi.org/10.1016/j.measurement.2019.02.072`.

[132] Niu L, Cao H, Hou H, Wu B, Lan Y, Xiong X. Experimental observations and dynamic modeling of vibration characteristics of a cylindrical roller bearing with roller defects. Mechanical Systems and Signal Processing. 2020;138:106553. Available from: `https://doi.org/10.1016/j.ymssp.2019.106553`.

[133] Singh J, Darpe AK, Singh SP. Rolling element bearing fault diagnosis based on Over-Complete rational dilation wavelet transform and auto-correlation of analytic energy operator. Mechanical Systems and Signal Processing. 2018;100:662-93. Available from: `http://dx.doi.org/10.1016/j.ymssp.2017.06.040`.

[134] Chen A, Kurfess TR. A new model for rolling element bearing defect size estimation. Measurement: Journal of the International Measurement Confederation. 2018;114(May 2017):144-9. Available from: `https://doi.org/10.1016/j.measurement.2017.09.018`.

[135] Zhang K, Lu Y, Shi H, Bai X, Yuan Z. The fault position identification method for the outer ring of deep groove ball bearing in a wide range. Measurement: Journal of the International Measurement Confederation. 2022;203(June 2021):111939. Available from: `https://doi.org/10.1016/j.measurement.2022.111939`.

[136] Zhang H, Borghesani P, Randall RB, Peng Z. A benchmark of measurement approaches to track the natural evolution of spall severity in rolling element bearings. Mechanical Systems and Signal Processing. 2022;166(July 2021):108466. Available from: `https://doi.org/10.1016/j.ymssp.2021.108466`.

[137] Guo Y, Sun SB, Wu X, Na J, Fung RF. Experimental investigation on double-impulse phenomenon of hybrid ceramic ball bearing with outer race spall. Mechanical Systems and Signal Processing. 2018;113:189-98. Available from: `https://doi.org/10.1016/j.ymssp.2016.07.042`.

[138] Chu WL, Lin CJ, Kao KC. Fault Diagnosis of a Rotor and Ball-Bearing System Using DWT Integrated with SVM, GRNN, and Visual Dot Patterns. Sensors. 2019;19:4806.

[139] Jiang Y, Huang W, Wang W, Peng G. A complete dynamics model of defective bearings considering the three-dimensional defect area and the spherical cage pocket. Mechanical Systems and Signal Processing. 2023;185(September 2022):109743. Available from: `https://doi.org/10.1016/j.ymssp.2022.109743`.

[140] Rohani Bastami A, Vahid S. A comprehensive evaluation of the effect of defect size in rolling element bearings on the statistical features of the vibration signal. Mechanical Systems and Signal Processing. 2021;151:107334. Available from: `https://doi.org/10.1016/j.ymssp.2020.107334`.

[141] Li Sy, Gu KR. Smart Fault-Detection Machine for Ball-Bearing System with Chaotic Mapping Strategy. Sensors (Switzerland). 2019;19:2178.

[142] Cui L, Huang J, Zhang F, Chu F. HVSRMS localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault. Mechanical Systems and Signal Processing. 2019;120:608-29. Available from: `https://doi.org/10.1016/j.ymssp.2018.09.043`.

[143] Liu Y, Han J, Zhao S, Meng Q, Shi T, Ma H. Study on the Dynamic Problems of Double-Disk Rotor System Supported by Deep Groove Ball Bearing. Shock and Vibration. 2019;2019.

[144] Han CF, Wu CC, Chu HY, Tsai PH, Horng JH, Wei CC, et al. Uses of empirical mode decomposition and multi-entropy techniques to establish the correlations among vibrations, friction coefficients and component wear of ball-bearing-like specimens. Measurement: Journal

of the International Measurement Confederation. 2020;150:107021. Available from: `https://doi.org/10.1016/j.measurement.2019.107021`.

[145] Dybała J. Diagnosing of rolling-element bearings using amplitude level-based decomposition of machine vibration signal. Measurement: Journal of the International Measurement Confederation. 2018;126(May 2017):143-55.

[146] Wang P, Yang Y, Ma H, Xu H, Li X, Luo Z, et al. Vibration characteristics of rotor-bearing system with angular misalignment and cage fracture: Simulation and experiment. Mechanical Systems and Signal Processing. 2023;182(February 2022).

[147] Zhang H, Borghesani P, Smith WA, Randall RB, Shahriar MR, Peng Z. Tracking the natural evolution of bearing spall size using cyclic natural frequency perturbations in vibration signals. Mechanical Systems and Signal Processing. 2021;151:107376. Available from: `https://doi.org/10.1016/j.ymssp.2020.107376`.

[148] Zhen D, Guo J, Xu Y, Zhang H, Gu F. A novel fault detection method for rolling bearings based on non-stationary vibration signature analysis. Sensors (Switzerland). 2019;19(18).

[149] Klausen A, Robbersmyr KG. Cross-correlation of whitened vibration signals for low-speed bearing diagnostics. Mechanical Systems and Signal Processing. 2019;118:226-44. Available from: `https://doi.org/10.1016/j.ymssp.2018.08.048`.

[150] Wu R, Wang X, Ni Z, Zeng C. Dual-impulse behavior analysis and quantitative diagnosis of the raceway fault of rolling bearing. Mechanical Systems and Signal Processing. 2022;169(July 2021):108734. Available from: `https://doi.org/10.1016/j.ymssp.2021.108734`.

[151] Jin Y, Liu Z, Yang Y, Li F, Chen Y. Nonlinear vibrations of a dual-rotor-bearing-coupling misalignment system with blade-casing rubbing. Journal of Sound and Vibration. 2021;497:22-37.

[152] Tian AH, Fu CB, Li YC, Yau HT. Intelligent ball bearing fault diagnosis using fractional Lorenz Chaos extension detection. Sensors (Switzerland). 2018;18(9).

[153] Kogan G, Klein R, Bortman J. A physics-based algorithm for the estimation of bearing spall width using vibrations. Mechanical Systems and Signal Processing. 2018;104:398-414. Available from: `https://doi.org/10.1016/j.ymssp.2017.11.011`.

[154] Zhang X, Yan C, Liu Y, Yan P, Wang Y, Wu L. Dynamic Modeling and Analysis of Rolling Bearing with Compound Fault on Raceway and Rolling Element. Shock and Vibration. 2020;2020.

[155] Jiang K, Li L, Han L, Gou S. In-Process Quality Inspection of Rolling Element Bearings Based on the Measurement of Microelastic Deformation of Outer Ring. Shock and Vibration. 2019;2019.

[156] Chen P, Li Y, Wang K, Zuo MJ. A novel knowledge transfer network with fluctuating operational condition adaptation for bearing fault pattern recognition. Measurement: Journal of the International Measurement Confederation. 2020;158.

[157] Jadhav PM, Kumbhar SG, Desavale RG, Patil SB. Distributed fault diagnosis of rotor-bearing system using dimensional analysis and experimental methods. Measurement: Journal of the International Measurement Confederation. 2020;166:1-12.

[158] Cheng Y, Chen B, Mei G, Wang Z, Zhang W. A novel blind deconvolution method and its application to fault identification. Journal of Sound and Vibration. 2019;460:114900. Available from: `https://doi.org/10.1016/j.jsv.2019.114900`.

[159] Rostaghi M, Ashory MR, Azami H. Application of dispersion entropy to status characterization of rotary machines. Journal of Sound and Vibration. 2019;438:291-308. Available from: https://doi.org/10.1016/j.jsv.2018.08.025.

[160] Li Y, Liang X, Liu W, Wang Y. Development of a morphological convolution operator for bearing fault detection. Journal of Sound and Vibration. 2018;421:220-33. Available from: https://doi.org/10.1016/j.jsv.2018.02.014.

[161] Cui L, Jin Z, Huang J, Wang H. Fault severity classification and size estimation for ball bearings based on vibration mechanism. IEEE Access. 2019;7:56107-16.

[162] Pham MT, Kim JM, Kim CH. Deep learning-based bearing fault diagnosis method for embedded systems. Sensors (Switzerland). 2020;20(23):1-15.

[163] Kong F, Huang W, Jiang Y, Wang W, Zhao X. A vibration model of ball bearings with a localized defect based on the hertzian contact stress distribution. Shock and Vibration. 2018;2018:1-14.

[164] Zhang B, Miao Y, Lin J, Liu Z. A new two-stage strategy to adaptively design and finely tune the filters for bearing fault-related mode decomposition. Measurement. 2023;210(May 2022):112470. Available from: https://doi.org/10.1016/j.measurement.2023.112470.

[165] Rohani Bastami A, Vahid S. Estimating the size of naturally generated defects in the outer ring and roller of a tapered roller bearing based on autoregressive model combined with envelope analysis and discrete wavelet transform. Measurement: Journal of the International Measurement Confederation. 2020;159:107767. Available from: https://doi.org/10.1016/j.measurement.2020.107767.

[166] Patil SM, Malagi RR, Desavale RG, Sawant SH. Fault identification in a nonlinear rotating system using Dimensional Analysis (DA) and central composite rotatable design (CCRD). Measurement: Journal of the International Measurement Confederation. 2022;200(May):111610. Available from: https://doi.org/10.1016/j.measurement.2022.111610.

[167] Hu C, Wang Y, Bai T. A Tensor-Based Approach for Identification of Multi-Channel Bearing Compound Faults. IEEE Access. 2019;7:38213-23.

[168] Nikula RP, Karioja K, Pylvänäinen M, Leiviskä K. Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features. Mechanical Systems and Signal Processing. 2020;138:106572. Available from: https://doi.org/10.1016/j.ymssp.2019.106572.

[169] Guo S, Bai C. Experimental and Numerical Research Studies on Multifield Coupling Rotor-dynamic Behaviors of Motor. Shock and Vibration. 2021;2021.

[170] Buchaiah S, Shakya P. Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. Measurement: Journal of the International Measurement Confederation. 2022;188(August 2021):110506. Available from: https://doi.org/10.1016/j.measurement.2021.110506.

[171] Shi HT, Bai XT. Model-based uneven loading condition monitoring of full ceramic ball bearings in starved lubrication. Mechanical Systems and Signal Processing. 2020;139:106583. Available from: https://doi.org/10.1016/j.ymssp.2019.106583.

[172] Chen Z, Mauricio A, Li W, Gryllias K. A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks. Mechanical Systems and Signal Processing. 2020;140:106683. Available from: https://doi.org/10.1016/j.ymssp.2020.106683.

[173] Luo Y, Chen C, Zhao S, Kong X, Wang Z. Rolling Bearing Diagnosis Based on Adaptive Probabilistic PCA and the Enhanced Morphological Filter. Shock and Vibration. 2020;2020.

[174] Liu S, Hou S, He K, Yang W. L-Kurtosis and its application for fault detection of rolling element bearings. Measurement: Journal of the International Measurement Confederation. 2018;116(November 2017):523-32. Available from: `https://doi.org/10.1016/j.measurement.2017.11.049`.

[175] Wang Y, Gu L, Qing L, Xu X, Shen J. An Improved k-Value Symmetrical Difference Analytic Energy Operator with HTFIF and L-KCA for Bearing Fault Diagnosis. IEEE Access. 2021;9:34307-24.

[176] Xin G, Hamzaoui N, Antoni J. Semi-automated diagnosis of bearing faults based on a hidden Markov model of the vibration signals. Measurement: Journal of the International Measurement Confederation. 2018;127(May):141-66. Available from: `https://doi.org/10.1016/j.measurement.2018.05.040`.

[177] Klausen A, Khang HV, Robbersmyr KG. Multi-band identification for enhancing bearing fault detection in variable speed conditions. Mechanical Systems and Signal Processing. 2020;139:106422. Available from: `https://doi.org/10.1016/j.ymssp.2019.106422`.

[178] Jiang Y, Huang W, Luo J, Wang W. An improved dynamic model of defective bearings considering the three-dimensional geometric relationship between the rolling element and defect area. Mechanical Systems and Signal Processing. 2019;129:694-716. Available from: `https://doi.org/10.1016/j.ymssp.2019.04.056`.

[179] Cui L, Wang X, Wang H, Wu N. Improved Fault Size Estimation Method for Rolling Element Bearings Based on Concatenation Dictionary. IEEE Access. 2019;7:22710-8.

[180] Li Q, Chen B, Zhang W, Song D. Research on mathematical morphological operators for fault diagnosis of rolling element bearings. Measurement: Journal of the International Measurement Confederation. 2022;203(September 2020):111964. Available from: `https://doi.org/10.1016/j.measurement.2022.111964`.

[181] Buzzoni M, Soave E, D'Elia G, Mucchi E, Dalpiaz G. Development of an Indicator for the Assessment of Damage Level in Rolling Element Bearings Based on Blind Deconvolution Methods. Shock and Vibration. 2018;2018.

[182] Yasir MN, Koh BH. Data decomposition techniques with multi-scale permutation entropy calculations for bearing fault diagnosis. Sensors (Switzerland). 2018;18(4).

[183] Gałęzia A, Gryllias K. Application of the Combined Teager-Kaiser Envelope for bearing fault diagnosis. Measurement: Journal of the International Measurement Confederation. 2021;182(June).

[184] Gu X, Yang S, Liu Y, Hao R, Liu Z. Multi-objective Informative Frequency Band Selection Based on Negentropy-induced Grey Wolf Optimizer for Fault Diagnosis of Rolling Element Bearings. Sensors. 2020;20:1845.

[185] Piltan F, Kim JM. Bearing fault diagnosis by a robust higher-order super-twisting sliding mode observer. Sensors (Switzerland). 2018;18(4):1-22.

[186] Udmale SS, Singh SK, Bhirud SG. A bearing data analysis based on kurtogram and deep learning sequence models. Measurement: Journal of the International Measurement Confederation. 2019;145:665-77. Available from: `https://doi.org/10.1016/j.measurement.2019.05.039`.

[187] Wang S, Niu P, Guo Y, Wang F, Li W, Shi H, et al. Early diagnosis of bearing faults using decomposition and reconstruction stochastic resonance system. Measurement: Journal of the International Measurement Confederation. 2020;158:107709. Available from: https://doi.org/10.1016/j.measurement.2020.107709.

[188] Shi H, Guo J, Yuan Z, Liu Z, Hou M, Sun J. Incipient Fault Detection of Rolling Element Bearings Based on Deep EMD-PCA Algorithm. Shock and Vibration. 2020;2020.

[189] Xu Y, Zhen D, Gu JX, Rabeyee K, Chu F, Gu F, et al. Autocorrelated Envelopes for early fault detection of rolling bearings. Mechanical Systems and Signal Processing. 2020;146:106990. Available from: https://doi.org/10.1016/j.ymssp.2020.106990.

[190] Jiang H, Yuan J, Zhao Q, Yan H, Wang S, Shao Y. A Robust Performance Degradation Modeling Approach Based on Student's t-HMM and Nuisance Attribute Projection. IEEE Access. 2020;8:49629-44.

[191] Buzzoni M, D'Elia G, Cocconcelli M. A tool for validating and benchmarking signal processing techniques applied to machine diagnosis. Mechanical Systems and Signal Processing. 2020;139:106618. Available from: https://doi.org/10.1016/j.ymssp.2020.106618.

[192] Liu H, Xiang J. A Strategy Using Variational Mode Decomposition, L-Kurtosis and Minimum Entropy Deconvolution to Detect Mechanical Faults. IEEE Access. 2019;7:70564-73.

[193] Yan X, Jia M. Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings. Mechanical Systems and Signal Processing. 2019;122:56-86. Available from: https://doi.org/10.1016/j.ymssp.2018.12.022.

[194] Xu L, Chatterton S, Pennacchi P. A novel method of frequency band selection for squared envelope analysis for fault diagnosing of rolling element bearings in a locomotive powertrain. Sensors (Switzerland). 2018;18(12).

[195] Rao M, Zuo MJ, Tian Z. A speed normalized autoencoder for rotating machinery fault detection under varying speed conditions. Mechanical Systems and Signal Processing. 2023;189(November 2022):110109. Available from: https://doi.org/10.1016/j.ymssp.2023.110109.

[196] Zhou S, Qian S, Chang W, Xiao Y, Cheng Y. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. Sensors (Switzerland). 2018;18(6).

[197] Wang D, Tsui KL, Qin Y. Optimization of segmentation fragments in empirical wavelet transform and its applications to extracting industrial bearing fault features. Measurement: Journal of the International Measurement Confederation. 2019;133:328-40. Available from: https://doi.org/10.1016/j.measurement.2018.10.018.

[198] Fonseca CA, Santos I, Weber HI. An experimental and theoretical approach of a pinned and a conventional ball bearing for active magnetic bearings. Mechanical Systems and Signal Processing. 2020;138:106541. Available from: https://doi.org/10.1016/j.ymssp.2019.106541.

[199] He Q, Wu E, Pan Y. Multi-Scale Stochastic Resonance Spectrogram for fault diagnosis of rolling element bearings. Journal of Sound and Vibration. 2018;420:174-84. Available from: https://doi.org/10.1016/j.jsv.2018.01.001.

[200] Yan X, Liu Y, Jia M. Research on an enhanced scale morphological-hat product filtering in incipient fault detection of rolling element bearings. Measurement: Journal of the International Measurement Confederation. 2019;147:106856. Available from: `https://doi.org/10.1016/j.measurement.2019.106856`.

[201] Hasan MJ, Sohaib M, Kim JM. A multitask-aided transfer learning-based diagnostic framework for bearings under inconsistent working conditions. Sensors (Switzerland). 2020;20(24):1-24.

[202] Zhang WT, Ji XF, Huang J, Lou ST. Compound Fault Diagnosis of Aero-Engine Rolling Element Bearing Based on CCA Blind Extraction. IEEE Access. 2021;9:159873-81.

[203] Duan J, Shi T, Zhou H, Xuan J, Zhang Y. Multiband envelope spectra extraction for fault diagnosis of rolling element bearings. Sensors (Switzerland). 2018;18(5):1-21.

[204] Ravikumar S, Kanagasabapathy H, Muralidharan V. Fault diagnosis of self-aligning troughing rollers in belt conveyor system using k-star algorithm. Measurement: Journal of the International Measurement Confederation. 2019;133:341-9. Available from: `https://doi.org/10.1016/j.measurement.2018.10.001`.

[205] Park J, Kim S, Choi JH, Lee SH. Frequency energy shift method for bearing fault prognosis using microphone sensor. Mechanical Systems and Signal Processing. 2021;147:107068. Available from: `https://doi.org/10.1016/j.ymssp.2020.107068`.

[206] Qin Y, Cao F, Wang Y, Chen W, Chen H. Dynamics modelling for deep groove ball bearings with local faults based on coupled and segmented displacement excitation. Journal of Sound and Vibration. 2019;447:1-19. Available from: `https://doi.org/10.1016/j.jsv.2019.01.048`.

[207] Gouda B, Tandon N, Pandey RK, Babu CK. Effect of grease grades on the tribological and vibrational behaviours of a radial ball bearing: Conventional vs. micro-grooved outer races. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. 2023;237(2):300-17.

[208] Pandey S, Amarnath M. Applications of vibro-acoustic measurement and analysis in conjunction with tribological parameters to assess surface fatigue wear developed in the roller-bearing system. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. 2021;235(10):2034-55.

[209] Kumar K, Shukla S, Singh SK. A combined approach for weak fault signature extraction of rolling element bearing using Hilbert envelop and zero frequency resonator. Journal of Sound and Vibration. 2018;419:436-51. Available from: `https://doi.org/10.1016/j.jsv.2018.01.022`.

[210] Shenfield A, Howarth M. A novel deep learning model for the detection and identification of rolling element-bearing faults. Sensors (Switzerland). 2020;20(18):1-24.

[211] Noman K, Wang D, Peng Z, He Q. Oscillation based permutation entropy calculation as a dynamic nonlinear feature for health monitoring of rolling element bearing. Measurement: Journal of the International Measurement Confederation. 2021;172(December 2020):108891. Available from: `https://doi.org/10.1016/j.measurement.2020.108891`.

[212] Gong T, Yuan X, Yuan Y, Lei X, Wang X. Application of tentative variational mode decomposition in fault feature detection of rolling element bearing. Measurement: Journal of the International Measurement Confederation. 2019;135:481-92. Available from: `https://doi.org/10.1016/j.measurement.2018.11.083`.

[213] Lv M, Liu S, Su X, Chen C. General Log-Linear Weibull Model Combining Vibration and Temperature Characteristics for Remaining Useful Life Prediction of Rolling Element Bearings. Shock and Vibration. 2020;2020.

[214] Yang Y, Fu P. Rolling-Element Bearing Fault Data Automatic Clustering Based on Wavelet and Deep Neural Network. Shock and Vibration. 2018;2018.

[215] Lin CJ, Su XY, Yu KT, Jian BL, Yau HT. Inspection on Ball Bearing Malfunction by Chen-Lee Chaos System. IEEE Access. 2020;8:28267-75.

[216] Song Q, Zhao S, Wang M. On the Accuracy of Fault Diagnosis for Rolling Element Bearings Using Improved DFA and Multi-Sensor Data Fusion Method. Sensors. 2020;20:6465.

[217] Wen X, Lu G, Liu J, Yan P. Graph modeling of singular values for early fault detection and diagnosis of rolling element bearings. Mechanical Systems and Signal Processing. 2020;145:106956. Available from: https://doi.org/10.1016/j.ymssp.2020.106956.

[218] Li L, Cui Y, Chen R, Liu X. Optimal SES Selection Based on SVD and Its Application to Incipient Bearing Fault Diagnosis. Shock and Vibration. 2018;2018.

[219] Li L, Cui Y, Chen R, Chen L, Wang L. Research on Rolling Bearing Fault Feature Extraction Based on Singular Value Decomposition considering the Singular Component Accumulative Effect and Teager Energy Operator. Shock and Vibration. 2019;2019(Vmd).

[220] Ni Q, Ji JC, Feng K, Halkon B. A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis. Mechanical Systems and Signal Processing. 2022;164(July 2021).

[221] Moshrefzadeh A, Fasana A. The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis. Mechanical Systems and Signal Processing. 2018;105:294-318. Available from: https://doi.org/10.1016/j.ymssp.2017.12.009.

[222] Sun H, Gao S, Ma S, Lin S. A fault mechanism-based model for bearing fault diagnosis under non-stationary conditions without target condition samples. Measurement: Journal of the International Measurement Confederation. 2022;199(March):111499. Available from: https://doi.org/10.1016/j.measurement.2022.111499.

[223] Jiao W, Li G, Jiang Y, Baim R, Tang C, Yan T, et al. Multi-Scale Sample Entropy-Based Energy Moment Features Applied to Fault Classification. IEEE Access. 2021;9:8444-54.

[224] Liang M, Su D, Hu D, Ge M. A Novel Faults Diagnosis Method for Rolling Element Bearings Based on ELCD and Extreme Learning Machine. Shock and Vibration. 2018;2018.

[225] Piltan F, Kim JM. Bearing fault diagnosis using an extended variable structure feedback linearization observer. Sensors (Switzerland). 2018;18(12):11-5.

[226] Rai A, Kim JM. A novel health indicator based on the Lyapunov exponent, a probabilistic self-organizing map, and the Gini-Simpson index for calculating the RUL of bearings. Measurement: Journal of the International Measurement Confederation. 2020;164:108002. Available from: https://doi.org/10.1016/j.measurement.2020.108002.

[227] Xiao L, Liu Z, Zhang Y, Zheng Y, Cheng C. Degradation assessment of bearings with trend-reconstruct-based features selection and gated recurrent unit network. Measurement: Journal of the International Measurement Confederation. 2020;165:108064. Available from: https://doi.org/10.1016/j.measurement.2020.108064.

[228] Duong BP, Khan SA, Shon D, Im K, Park J, Lim DS, et al. A reliable health indicator for fault prognosis of bearings. Sensors (Switzerland). 2018;18(11).

[229] Kass S, Raad A, Antoni J. Self-running bearing diagnosis based on scalar indicator using fast order frequency spectral coherence. Measurement: Journal of the International Measurement Confederation. 2019;138:467-84. Available from: https://doi.org/10.1016/j.measurement.2019.02.046.

[230] Kundu P, Darpe AK, Kulkarni MS. Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions. Mechanical Systems and Signal Processing. 2019;134:106302. Available from: https://doi.org/10.1016/j.ymssp.2019.106302.

[231] Kannan V, Li H, Dao DV. Demodulation Band Optimization in Envelope Analysis for Fault Diagnosis of Rolling Element Bearings Using a Real-Coded Genetic Algorithm. IEEE Access. 2019;7:168828-38.

[232] Wang B, Lei Y, Li N, Yan T. Deep separable convolutional network for remaining useful life prediction of machinery. Mechanical Systems and Signal Processing. 2019;134:106330. Available from: https://doi.org/10.1016/j.ymssp.2019.106330.

[233] Liu S, Sun Y, Zhang L. A Novel Fault Diagnosis Method Based on Noise-Assisted MEMD and Functional Neural Fuzzy Network for Rolling Element Bearings. IEEE Access. 2018;6:27048-68.

[234] Chen A, Kurfess TR. Signal processing techniques for rolling element bearing spall size estimation. Mechanical Systems and Signal Processing. 2019;117:16-32. Available from: https://doi.org/10.1016/j.ymssp.2018.03.006.

[235] Tiwari P, Upadhyay SH. Novel self-adaptive vibration signal analysis: Concealed component decomposition and its application in bearing fault diagnosis. Journal of Sound and Vibration. 2021;502:116079. Available from: https://doi.org/10.1016/j.jsv.2021.116079.

[236] Bai X, Zheng H, Wang Z, Wang Z. Raceway Defect Frequency Deviation of Full-Ceramic Ball Bearing Induced by Fit Clearance in Wide Temperature Ranges. Shock and Vibration. 2021;2021.

[237] Li Y, Zuo MJ, Chen Y, Feng K. An enhanced morphology gradient product filter for bearing fault detection. Mechanical Systems and Signal Processing. 2018;109:166-84. Available from: https://doi.org/10.1016/j.ymssp.2018.03.002.

[238] Gong T, Yuan X, Lei X, Yuan Y, Zhang B. Fault detection for rolling element bearing based on repeated single-scale morphology and simplified sensitive factor algorithm. Measurement: Journal of the International Measurement Confederation. 2018;127(May):348-55. Available from: https://doi.org/10.1016/j.measurement.2018.05.082.

[239] Xiao L, Zhang X, Lu S, Xia T, Xi L. A novel weak-fault detection technique for rolling element bearing based on vibrational resonance. Journal of Sound and Vibration. 2019;438:490-505. Available from: https://doi.org/10.1016/j.jsv.2018.09.039.

[240] Xu M, Han Y, Sun X, Shao Y, Gu F, Ball AD. Vibration characteristics and condition monitoring of internal radial clearance within a ball bearing in a gear-shaft-bearing system. Mechanical Systems and Signal Processing. 2022;165(July 2021):108280. Available from: https://doi.org/10.1016/j.ymssp.2021.108280.

[241] Guo J, Zhang H, Zhen D, Shi Z, Gu F, Ball AD. An enhanced modulation signal bispectrum analysis for bearing fault detection based on non-Gaussian noise suppression. Measurement: Journal of the International Measurement Confederation. 2020;151:107240. Available from: `https://doi.org/10.1016/j.measurement.2019.107240`.

[242] Berntsen J, Brandt A, Gryllias K. Enhanced demodulation band selection based on Operational Modal Analysis (OMA) for bearing diagnostics. Mechanical Systems and Signal Processing. 2022;181(April):109300. Available from: `https://doi.org/10.1016/j.ymssp.2022.109300`.

[243] Yang T, Guo Y, Wu X, Na J, Fung RF. Fault feature extraction based on combination of envelope order tracking and cICA for rolling element bearings. Mechanical Systems and Signal Processing. 2018;113:131-44. Available from: `https://doi.org/10.1016/j.ymssp.2017.03.050`.

[244] He Y, Feng K, Hu M, Cui J. An MCM-Enhanced Compressive Sensing for Weak Fault Feature Extraction of Rolling Element Bearings under Variable Speeds. Shock and Vibration. 2020;2020.

[245] Mauricio A, Qi J, Smith WA, Sarazin M, Randall RB, Janssens K, et al. Bearing diagnostics under strong electromagnetic interference based on Integrated Spectral Coherence. Mechanical Systems and Signal Processing. 2020;140:106673. Available from: `https://doi.org/10.1016/j.ymssp.2020.106673`.

[246] Cheng Y, Zhou N, Zhang W, Wang Z. Application of an improved minimum entropy deconvolution method for railway rolling element bearing fault diagnosis. Journal of Sound and Vibration. 2018;425:53-69. Available from: `https://doi.org/10.1016/j.jsv.2018.01.023`.

[247] Xu G, Hou D, Qi H, Bo L. High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life. Mechanical Systems and Signal Processing. 2021;146:107050. Available from: `https://doi.org/10.1016/j.ymssp.2020.107050`.

[248] Huang Y, Lin J, Liu Z, Wu W. A modified scale-space guiding variational mode decomposition for high-speed railway bearing fault diagnosis. Journal of Sound and Vibration. 2019;444:216-34. Available from: `https://doi.org/10.1016/j.jsv.2018.12.033`.

[249] Zhang B, Miao Y, Lin J, Yi Y. Adaptive maximum second-order cyclostationarity blind deconvolution and its application for locomotive bearing fault diagnosis. Mechanical Systems and Signal Processing. 2021;158:107736. Available from: `https://doi.org/10.1016/j.ymssp.2021.107736`.

[250] Luo Y, Chen C, Zhao S, Yang G. Rolling bearing fault diagnosis method with enhanced top-hat transform filtering and cyclic spectrum coherence. IEEE Access. 2020;8:163715-29.

[251] Huang X, Wen G, Liang L, Zhang Z, Tan Y. Frequency Phase Space Empirical Wavelet Transform for Rolling Bearings Fault Diagnosis. IEEE Access. 2019;7:86306-18.

[252] Zheng Z, Li X, Zhu Y. Feature extraction of the hydraulic pump fault based on improved Autogram. Measurement: Journal of the International Measurement Confederation. 2020;163:107908. Available from: `https://doi.org/10.1016/j.measurement.2020.107908`.

[253] Wang S, Jing B, Jiao X, Pan J. Degradation Pattern Identification and Remaining Useful Life Prediction for Mechanical Equipment Using SKF-EN. IEEE Access. 2020;8:147662-72.

[254] Yang B, Lei Y, Jia F, Xing S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mechanical Systems and Signal Processing. 2019;122:692-706. Available from: https://doi.org/10.1016/j.ymssp.2018.12.051.

[255] Liu Z, Ding J, Lin J, Huang Y. A rolling bearing fault diagnosis-optimized scale-space representation for the empirical wavelet transform. Shock and Vibration. 2018;2018.

[256] Kim SJ, Kim K, Hwang T, Park J, Jeong H, Kim T, et al. Motor-current-based electromagnetic interference de-noising method for rolling element bearing diagnosis using acoustic emission sensors. Measurement: Journal of the International Measurement Confederation. 2022;193(October 2021):110912. Available from: https://doi.org/10.1016/j.measurement.2022.110912.

[257] Qiu C, Wu X, Xu C, Qiu X, Xue Z. An approximate estimation approach of fault size for spalled ball bearing in induction motor by tracking multiple vibration frequencies in current. Sensors (Switzerland). 2020;20(6).

[258] Jablonka K, Glovnea R, Bongaerts J. Quantitative measurements of film thickness in a radially loaded deep-groove ball bearing. Tribology International. 2018;119(November 2017):239-49. Available from: https://doi.org/10.1016/j.triboint.2017.11.001.

[259] Han Q, Ding Z, Qin Z, Wang T, Xu X, Chu F. A triboelectric rolling ball bearing with self-powering and self-sensing capabilities. Nano Energy. 2020;67(September 2019):104277. Available from: https://doi.org/10.1016/j.nanoen.2019.104277.

[260] Bai X, Shi H, Zhang K, Zhang X, Wu Y. Effect of the fit clearance between ceramic outer ring and steel pedestal on the sound radiation of full ceramic ball bearing system. Journal of Sound and Vibration. 2022;529(August 2021):116967. Available from: https://doi.org/10.1016/j.jsv.2022.116967.

[261] Luo M, Guo Y, Wu X, Na J. An analytical model for estimating spalled zone size of rolling element bearing based on dual-impulse time separation. Journal of Sound and Vibration. 2019;453:87-102. Available from: https://doi.org/10.1016/j.jsv.2019.04.014.

[262] Saufi SR, Ahmad ZAB, Leong MS, Lim MH. Low-Speed Bearing Fault Diagnosis Based on ArSSAE Model Using Acoustic Emission and Vibration Signals. IEEE Access. 2019;7:46885-97.

[263] Li Z, Zhao C, Lu Z, Liu F. Thermal performances prediction analysis of high speed feed shaft bearings under actual working condition. IEEE Access. 2019;7:168011-9.

[264] Brossier P, Niel D, Changenet C, Ville F, Belmonte J. Experimental and numerical investigations on rolling element bearing thermal behaviour. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. 2021;235(4):842-53.

[265] Ohta H, Oguma K, Takane K, Kato S. Friction Forces in a Linear-Guideway Type Recirculating Ball Bearing under Grease Lubrication. Journal of Tribology. 2021;143(6):1-8.

[266] Yunlong W, Yulong L, Ziqiang Z, Wenzhong W. Lubrication and Thermal Failure Mechanism Analysis in High-Speed Angular Contact Ball Bearing. Journal of Tribology. 2018;140(3):1-11.

[267] Zhang Y, hu Li X, Hong J, Yan K, Li S. Uneven heat generation and thermal performance of spindle bearings. Tribology International. 2018;126(February):324-35. Available from: https://doi.org/10.1016/j.triboint.2018.04.035.

[268] Clarke BP, Nicholas G, Hart E, Long H, Dwyer-joyce RS. Loading on a wind turbine high-speed shaft gearbox bearing : Ultrasonic field measurements and predictions from a multi-body simulation. Tribology International. 2023;181(October 2022):108319. Available from: https://doi.org/10.1016/j.triboint.2023.108319.

[269] Sikder N, Bhakta K, Al Nahid A, Islam MMM. Fault diagnosis of motor bearing using ensemble learning algorithm with FFT-based preprocessing. In: 1st International Conference on Robotics, Electrical and Signal Processing Techniques, ICREST 2019. IEEE; 2019. p. 564-9.

[270] Tang S, Yuan S, Zhu Y. Data Preprocessing Techniques in Convolutional Neural Network Based on Fault Diagnosis towards Rotating Machinery. IEEE Access. 2020;8:149487-96.

[271] Feng GJ, Gu J, Zhen D, Aliwan M, Gu FS, Ball AD. Implementation of envelope analysis on a wireless condition monitoring system for bearing fault diagnosis. International Journal of Automation and Computing. 2015;12(1):14-24.

[272] Bingamil A, Alsyouf I, Cheaitou A. Condition monitoring technologies, parameters and data processing techniques for fault detection of internal combustion engines: A literature review. 2017 International Conference on Electrical and Computing Technologies and Applications, ICECTA 2017. 2017;2018-January:1-5.

[273] Markovitch S, Rosenstein D. Feature generation using general constructor functions. Machine Learning. 2002;49(1):59-98.

[274] Famili A, Shen WM, Weber R, Simoudis E. Data Preprocessing and Intelligent Data Analysis. Intelligent Data Analysis. 1997;1(1):3-23.

[275] Alasadi SA, Bhaya WS. Review of Data Preprocessing Techniques. Journal of Engineering and Applied Sciences. 2017;12(16):4102-7.

[276] Ramírez-Gallego S, Krawczyk B, García S, Woźniak M, Herrera F. A survey on data preprocessing for data stream mining: Current status and future directions. Neurocomputing. 2017;239:39-57.

[277] Alexandropoulos SAN, Kotsiantis SB, Vrahatis MN. Data preprocessing in predictive data mining. Knowledge Engineering Review. 2019;34:1-33.

[278] Han Y, Song YH. Condition monitoring techniques for electrical equipment - A literature survey. IEEE Transactions on Power Delivery. 2003;18(1):4-13.

[279] Nidadavolu SVPS, Yadav SK, Kalra PK. Condition monitoring of internal combustion engines using empirical mode decomposition and morlet wavelet. In: ISPA 2009 - Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis. IEEE; 2009. p. 75-82.

[280] Chochran WT, Cooley JW, Favin DL, Helms HD, Kaenel RA, Lang WW, et al. What Is the The fast Fourier transform. Proceedings of the IEEE. 1967;55(10):1664 1674.

[281] Hameed Z, Hong YS, Cho YM, Ahn SH, Song CK. Condition monitoring and fault detection of wind turbines and related algorithms: A review. Renewable and Sustainable Energy Reviews. 2009 jan;13(1):1-39.

[282] Rai VK, Mohanty AR. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform. Mechanical Systems and Signal Processing. 2007 aug;21(6):2607-15.

[283] BahooToroody A, Abaei MM, BahooToroody F, De Carlo F, Abbassi R, Khalaj S. A condition monitoring based signal filtering approach for dynamic time dependent safety assessment of natural gas distribution process. Process Safety and Environmental Protection. 2019;123:335-43. Available from: https://doi.org/10.1016/j.psep.2019.01.016.

[284] Chaudhari H, Nalbalwar SL, Sheth R. A review on intrensic mode function of EMD. In: International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016. IEEE; 2016. p. 2349-52.

[285] Torrence C, Compo GP. A Practical Guide to Wavelet Analysis. Bulletin of the American Meteorological Society. 1998;79(1):61-78.

[286] Grossmann A, Morlet J. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. J Math Anal. 1984;15(4):723 736. Available from: http://www.citeulike.org/group/7763/article/3716013.

[287] Yang W, Court R, Jiang J. Wind turbine condition monitoring by the approach of SCADA data analysis. Renewable Energy. 2013 may;53:365-76.

[288] Ompusunggu AP, Eryilmaz K, Janssen K. Condition monitoring of critical industrial assets using high performing low-cost MEMS accelerometers. Procedia CIRP. 2021;104:1389-94. Available from: https://doi.org/10.1016/j.procir.2021.11.234.

[289] Wu W, Lin TR, Tan ACC. Normalization and source separation of acoustic emission signals for condition monitoring and fault detection of multi-cylinder diesel engines. Mechanical Systems and Signal Processing. 2015;64-65:479-97. Available from: http://dx.doi.org/10.1016/j.ymssp.2015.03.016.

[290] Yang C, Liu J, Zeng Y, Xie G. Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model. Renewable Energy. 2019;133:433-41. Available from: https://doi.org/10.1016/j.renene.2018.10.062.

[291] Dyer D, Stewart RM. Detection of rolling element bearing damage by statistical vibration analysis. Journal of Mechanical Design, Transactions of the ASME. 1978;100(2):229-35.

[292] Mechefske CK. Machine condition monitoring and fault diagnostics. In: Vibration and Shock Handbook. Boca Raton: Taylor and Francis Group; 2007. p. 329-63.

[293] Helmi H, Forouzantabar A. Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS. IET Electric Power Applications. 2019;13(5):662-9.

[294] Kurfess TR, Billington S, Liang SY. Advanced Diagnostic and Prognostic Techniques for Rolling Element Bearings. In: Wang L, Gao RX, editors. Condition Monitoring and Control for Intelligent Manufacturing. Springer London; 2006. p. 137-66.

[295] Ren L, Cui J, Sun Y, Cheng X. Multi-bearing remaining useful life collaborative prediction: A deep learning approach. Journal of Manufacturing Systems. 2017 apr;43:248-56.

[296] Caesarendra W, Tjahjowidodo T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. Machines. 2017 dec;5(4).

[297] Nayana BR, Geethanjali P. Analysis of Statistical Time-Domain Features Effectiveness in Identification of Bearing Faults from Vibration Signal. IEEE Sensors Journal. 2017 sep;17(17):5618-25.

[298] Buhl WT. Ball Bearing Noise and Vibration. Journal of the Acoustical Society of America. 1941;13(1):51-3.

[299] Broderick JJ, Burchill RF, Clark HL. Design and Fabrication of Prototype System for Early Warning of Impending Bearing Failure. Huntsville: George C Marshall Space Flight Center, NASA; 1972.

[300] McFadden PD, Smith JD. Vibration monitoring of rolling element bearings by the high-frequency resonance technique - a review. Tribology International. 1984;17(1):3-10.

[301] McFadden PD, Smith JD. The vibration produced by multiple point defects in a rolling element bearing. Journal of Sound and Vibration. 1985;98(2):263-73.

[302] McFadden PD, Smith JD. Model for the vibration produced by a single point defect in a rolling element bearing. Journal of Sound and Vibration. 1984;96(1):69-82.

[303] Chen Z, Cen J, Xiong J. Rolling Bearing Fault Diagnosis Using Time-Frequency Analysis and Deep Transfer Convolutional Neural Network. IEEE Access. 2020;8:150248-61.

[304] Tian J, Morillo C, Azarian MH, Pecht M. Kurtosis-Based Feature Extraction Coupled With K -Nearest Neighbor Distance Analysis. IEEE Trans Ind Electron. 2016;63(3):1793-803.

[305] Wang J, Mo Z, Zhang H, Miao Q. A deep learning method for bearing fault diagnosis based on time-frequency image. IEEE Access. 2019;7:42373-83.

[306] Yang Y, Yu D, Cheng J. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. Measurement: Journal of the International Measurement Confederation. 2007;40(9-10):943-50.

[307] Cerrada M, Sánchez RV, Li C, Pacheco F, Cabrera D, Valente de Oliveira J, et al. A review on data-driven fault severity assessment in rolling bearings. Mechanical Systems and Signal Processing. 2018 jan;99:169-96.

[308] El-Thalji I, Jantunen E. A summary of fault modelling and predictive health monitoring of rolling element bearings. Mechanical Systems and Signal Processing. 2015;60:252-72. Available from: http://dx.doi.org/10.1016/j.ymssp.2015.02.008.

[309] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing. 2020;138:106587. Available from: https://doi.org/10.1016/j.ymssp.2019.106587.

[310] Liu R, Yang B, Zio E, Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing. 2018;108:33-47. Available from: https://doi.org/10.1016/j.ymssp.2018.02.016.

[311] Hoang DT, Kang HJ. A survey on Deep Learning based bearing fault diagnosis. Neurocomputing. 2019;335:327-35. Available from: https://doi.org/10.1016/j.neucom.2018.06.078.

[312] Lei Y, Lin J, Zuo MJ, He Z. Condition monitoring and fault diagnosis of planetary gearboxes: A review. Measurement: Journal of the International Measurement Confederation. 2014;48(1):292-305. Available from: http://dx.doi.org/10.1016/j.measurement.2013.11.012.

[313] Hamadache M, Jung JH, Park J, Youn BD. A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning. JMST Advances. 2019;1(1-2):125-51. Available from: https://doi.org/10.1007/s42791-019-0016-y.

[314] Peng Y, Dong M, Zuo MJ. Current status of machine prognostics in condition-based maintenance: A review. International Journal of Advanced Manufacturing Technology. 2010 sep;50(1-4):297-313.

[315] Zhang S, Zhang S, Wang B, Habetler TG. Deep Learning Algorithms for Bearing Fault Diagnostics - A Comprehensive Review. IEEE Access. 2020;8:29857-81.

[316] Peng B, Bi Y, Xue B, Zhang M, Wan S. A Survey on Fault Diagnosis of Rolling Bearings. Algorithms. 2022;15(10):1-24.

[317] Alshorman O, Irfan M, Saad N, Zhen D, Haider N, Glowacz A, et al. A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor. Shock and Vibration. 2020;2020(Cm).

[318] Kotsiantis SB. Supervised Machine Learning: A Review of Classification Techniques. Informatica. 2007;31(1):249 268.

[319] Maio FD, Ng SSY, Tsui KL, Zio E. Naïve Bayesian Classifier for On-line Remaining Useful Life Prediction of Degrading Bearings. In: Proceedings of the Seventh International Conference on Mathematical Methods in Reliability. June; 2011. p. 1-14.

[320] Toma RN, Prosvirin AE, Kim JM. Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. Sensors. 2020;20(7):1 19.

[321] Paya BA, Esat II, Badi MNM. Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. Mechanical Systems and Signal Processing. 1997;11(5):751-65.

[322] Hu J, Zou W, Wang J, Pang L. Minimum training sample size requirements for achieving high prediction accuracy with the BN model: A case study regarding seismic liquefaction. Expert Systems with Applications. 2021;185(July):115702. Available from: `https://doi.org/10.1016/j.eswa.2021.115702`.

[323] Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences of the United States of America. 2019;116(32):15849-54.

[324] Längkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters. 2014;42(1):11-24.

[325] Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. Neurocomputing. 2017;267:664-81. Available from: `http://dx.doi.org/10.1016/j.neucom.2017.06.053`.

[326] Al Farizi WS, Hidayah I, Rizal MN. Isolation Forest Based Anomaly Detection: A Systematic Literature Review. 2021 8th International Conference on Information Technology, Computer and Electrical Engineering, ICITACEE 2021. 2021:118-22.

[327] Alghushairy O, Alsini R, Soule T, Ma X. A review of local outlier factor algorithms for outlier detection in big data streams. Big Data and Cognitive Computing. 2021;5(1):1-24.

[328] Vakharia V, Gupta VK, Kankar PK. Ball Bearing Fault Diagnosis using Supervised and Unsupervised Machine Learning Methods. The International Journal of Acoustics and Vibration. 2015;20(4).

[329] Aljemely AH, Xuan J, Jawad FKJ, Al-Azzawi O, Alhumaima AS. A novel unsupervised learning method for intelligent fault diagnosis of rolling element bearings based on deep functional auto-encoder. Journal of Mechanical Science and Technology. 2020;34(11):4367-81.

[330] Li YF, Liang DM. Safe semi-supervised learning: a brief introduction. Frontiers of Computer Science. 2019;13(4):669-76.

[331] Gomes HM, Grzenda M, Mello R, Read J, Le Nguyen MH, Bifet A. A Survey on Semi-supervised Learning for Delayed Partially Labelled Data Streams. ACM Computing Surveys. 2022;55(4).

[332] Bull LA, Worden K, Dervilis N. Towards semi-supervised and probabilistic classification in structural health monitoring. Mechanical Systems and Signal Processing. 2020;140:106653. Available from: `https://doi.org/10.1016/j.ymssp.2020.106653`.

[333] Yuan J, Liu X. Semi-supervised learning and condition fusion for fault diagnosis. Mechanical Systems and Signal Processing. 2013;38(2):615-27. Available from: `http://dx.doi.org/10.1016/j.ymssp.2013.03.008`.

[334] ISO. ISO 281:2007 Rolling bearings - Dynamic load ratings and rating life. ISO; 2007.

[335] Tandon N, Choudhury A. An Analytical Model for the Prediction of the Vibration Response of Rolling Element Bearing Due to a Localized Defect; 1997. 3.

[336] Sassi S, Badri B, Thomas M. A numerical model to predict damaged bearing vibrations. JVC/Journal of Vibration and Control. 2007;13(11):1603-28.

[337] Liu J, Shao Y, Lim TC. Vibration analysis of ball bearings with a localized defect applying piecewise response function. Mechanism and Machine Theory. 2012 oct;56:156-69.

[338] Hertz H. On the contact of elastic solids. J Reine Agnew Math. 1881;92:156-71.

[339] Sawalhi N, Randall RB. Simulating gear and bearing interactions in the presence of faults. Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults. Mechanical Systems and Signal Processing. 2008;22(8):1924-51.

[340] Rycerz P, Olver A, Kadiric A. Propagation of surface initiated rolling contact fatigue cracks in bearing steel. International Journal of Fatigue. 2017;97:29-38. Available from: `http://dx.doi.org/10.1016/j.ijfatigue.2016.12.004`.

[341] Qian Y, Yan R, Gao RX. A multi-time scale approach to remaining useful life prediction in rolling bearing. Mechanical Systems and Signal Processing. 2017;83:549-67. Available from: `http://dx.doi.org/10.1016/j.ymssp.2016.06.031`.

[342] Yuan Z, Wang B, Liu C, Wang Z, Zhang X, Zhang Y. The Crack Propagation Trend Analysis in Ceramic Rolling Element Bearing considering Initial Crack Angle and Contact Load Effect. Shock and Vibration. 2021;2021.

[343] Forman RG, Kearney VE, Engle RM. Numerical analysis of crack propagation in cyclic-loaded structures. Journal of Fluids Engineering, Transactions of the ASME. 1967;89(3):459-63.

[344] Elber W. The Significance of Fatigue Crack Closure. In: Damge Tolerance in Aircraft Structures, ASTM STP 486. American Society for Testing and Materials; 1971. p. 230-42.

[345] Shahid AH, Singh MP. Computational intelligence techniques for medical diagnosis and prognosis: Problems and current developments. Biocybernetics and Biomedical Engineering. 2019;39(3):638-72. Available from: `https://doi.org/10.1016/j.bbe.2019.05.010`.

[346] Sadoughi M, Hu C. Physics-Based Convolutional Neural Network for Fault Diagnosis of Rolling Element Bearings. IEEE Sensors Journal. 2019 jun;19(11):4181-92.

[347] Tobon-Mejia DA, Medjaher K, Zerhouni N, Tripot G. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. IEEE Transactions on Reliability. 2012;61(2):491-503.

[348] Soave E, D'Elia G, Dalpiaz G. Prognostics of rotating machines through generalized Gaussian hidden Markov models. Mechanical Systems and Signal Processing. 2023;185(March 2022):109767. Available from: https://doi.org/10.1016/j.ymssp.2022.109767.

[349] Ni Q, Ji JC, Feng K. Data-Driven Prognostic Scheme for Bearings Based on a Novel Health Indicator and Gated Recurrent Unit Network. IEEE Transactions on Industrial Informatics. 2023;19(2):1301-11.

[350] Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Computing Surveys. 2019;52(4).

[351] Dyk DAV, Meng XL. The art of data augmentation. Journal of Computational and Graphical Statistics. 2001;10(1):1-50.

[352] Torrey L, Shavlik J. Transfer Learning. In: Soria E, Martin J, Magdalena R, Martinez M, Serrano A, editors. Handbook of Research on Machine Learning Applications. IGI Global; 2009. p. 1 22.

[353] Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter. 2004;6(1):20-9.

[354] Yijing L, Haixiang G, Xiao L, Yanan L, Jinling L. Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. Knowledge-Based Systems. 2016;94:88-104.

[355] Sowah RA, Agebure MA, Mills GA, Koumadi KM, Fiawoo SY. New Cluster Undersampling Technique for Class Imbalance Learning. International Journal of Machine Learning and Computing. 2016;6(3):205-14.

[356] Spiesman BJ, Gratton C, Hatfield RG, Hsu WH, Jepsen S, McCornack B, et al. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. Scientific Reports. 2021;11(1):1-10. Available from: https://doi.org/10.1038/s41598-021-87210-1.

[357] Castelvecchi D. Beating Biometric Bias. Nature. 2020 nov:347-9.

[358] Janai J, Güney F, Behl A, Geiger A. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. Foundations and Trends in Computer Graphics and Computer Vision. 2020;12(1-3):1-308.

[359] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. Journal of Big Data. 2019 dec;6(1).

[360] Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology. 2021;65(5):545-63.

[361] Liu P, Wang X, Xiang C, Meng W. A Survey of Text Data Augmentation. Proceedings - 2020 International Conference on Computer Communication and Network Security, CCNS 2020. 2020:191-5.

[362] Zhang W, Li C, Peng G, Chen Y, Zhang Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mechanical Systems and Signal Processing. 2018;100:439-53. Available from: `http://dx.doi.org/10.1016/j.ymssp.2017.06.022`.

[363] Li X, Zhang W, Ding Q, Sun JQ. Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. Journal of Intelligent Manufacturing. 2020 feb;31(2):433-52.

[364] Yu K, Lin TR, Ma H, Li X, Li X. A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. Mechanical Systems and Signal Processing. 2021 jan;146.

[365] Bui V, Pham TL, Nguyen H, Jang YM. Data augmentation using generative adversarial network for automatic machine fault detection based on vibration signals. Applied Sciences. 2021;11(5):1-16.

[366] Shao S, Wang P, Yan R. Generative adversarial networks for data augmentation in machine fault diagnosis. Computers in Industry. 2019 apr;106:85-93.

[367] Wang J, Han B, Bao H, Wang M, Chu Z, Shen Y. Data augment method for machine fault diagnosis using conditional generative adversarial networks. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering. 2020;234(12):2719-27.

[368] Schwendemann S, Amjad Z, Sikora A. Bearing fault diagnosis with intermediate domain based Layered Maximum Mean Discrepancy: A new transfer learning approach. Engineering Applications of Artificial Intelligence. 2021;105(July):104415. Available from: `https://doi.org/10.1016/j.engappai.2021.104415`.

[369] Zhu J, Chen N, Shen C. A New Deep Transfer Learning Method for Bearing Fault Diagnosis Under Different Working Conditions. IEEE Sensors Journal. 2020;20(15):8394-402.

[370] Zhao K, Jiang H, Wu Z, Lu T. A novel transfer learning fault diagnosis method based on Manifold Embedded Distribution Alignment with a little labeled data. Journal of Intelligent Manufacturing. 2022;33(1):151-65. Available from: `https://doi.org/10.1007/s10845-020-01657-z`.

[371] Chen J, Huang R, Chen Z, Mao W, Li W. Transfer learning algorithms for bearing remaining useful life prediction : A comprehensive review from an industrial application perspective. Mechanical Systems and Signal Processing. 2023;193(December 2022):110239. Available from: `https://doi.org/10.1016/j.ymssp.2023.110239`.

[372] Chen X, Yang R, Xue Y, Huang M. Deep Transfer Learning for Bearing Fault Diagnosis : A Systematic Review Since 2016. IEEE Transactions on Instrumentation and Measurement. 2023;72:1-21.

[373] Li W, Huang R, Li J, Liao Y, Chen Z, He G, et al. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. Mechanical Systems and Signal Processing. 2022;167(PA):108487. Available from: `https://doi.org/10.1016/j.ymssp.2021.108487`.

[374] Grieves M, Vickers J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In: Kahlen FJ, Flumerfelt S, Alves A, editors. Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. August. Spring, Cham.; 2017. p. 85 113.

[375] Shao G, Helu M. Framework for a digital twin in manufacturing: Scope and requirements. Manufacturing Letters. 2020;24:105-7. Available from: `https://doi.org/10.1016/j.mfglet.2020.04.004`.

[376] You Y, Chen C, Hu F, Liu Y, Ji Z. Advances of Digital Twins for Predictive Maintenance. Procedia Computer Science. 2022;200(2019):1471-80.

[377] Moghadam FK, Nejad AR. Online condition monitoring of floating wind turbines drivetrain by means of digital twin. Mechanical Systems and Signal Processing. 2022;162(December 2020):108087. Available from: `https://doi.org/10.1016/j.ymssp.2021.108087`.

[378] Qin Y, Wu X, Luo J. Data-Model Combined Driven Digital Twin of Life-Cycle Rolling Bearing. IEEE Transactions on Industrial Informatics. 2022;18(3):1530-40.

[379] Lee EA. Cyber physical systems: Design challenges. In: Proceedings - 11th IEEE Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing, ISORC 2008; 2008. p. 363-9.

[380] Tao F, Qi Q, Wang L, Nee AYC. Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. Engineering. 2019 aug;5(4):653-61.

[381] Fredriksson T, Mattos DI, Bosch J, Olsson HH. Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation Strategies. In: Product-Focused Software Process Improvement. March 2021; 2020. p. 202-16.

[382] Barbosa NM, Chen M. Rehumanized crowdsourcing: A labeling framework addressing bias and ethics in machine learning. In: Conference on Human Factors in Computing Systems - Proceedings. Association for Computing Machinery; 2019. p. 1-12.

[383] Chen LC, Fidler S, Yuille AL, Urtasun R. Beat the MTurkers: Automatic Image Labeling from Weak 3D Supervision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 3198-205. Available from: `https://bitbucket.org/liang`.

[384] Zhang J, Wu X, Sheng VS. Learning from crowdsourced labeled data: a survey. Artificial Intelligence Review. 2016;46(4):543-76.

[385] Jiang H, Nachum O. Identifying and Correcting Label Bias in Machine Learning. In: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. vol. PMLR 108; 2020. p. 702-12. Available from: `http://arxiv.org/abs/1901.04966`.

[386] Bangalore P, Tjernberg LB. An artificial neural network approach for early fault detection of gearbox bearings. IEEE Transactions on Smart Grid. 2015;6(2):980-7.

[387] Egusquiza E, Valero C, Valentin D, Presas A, Rodriguez CG. Condition monitoring of pump-turbines. New challenges. Measurement: Journal of the International Measurement Confederation. 2015;67:151-63. Available from: `http://dx.doi.org/10.1016/j.measurement.2015.01.004`.

[388] Liu Z, Zhang L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. Measurement. 2020 jan;149:1-22.

[389] ISO. ISO 20816-1:Mechanical vibration — Measurement and evaluation of machine vibration — Part 1: General guidelines. Geneva: International Organization for Standardization; 2016.

[390] Lyu Y, Chen J, Song Z. Synthesizing labeled data to enhance soft sensor performance in data-scarce regions. Control Engineering Practice. 2021;115(December 2020):104903. Available from: https://doi.org/10.1016/j.conengprac.2021.104903.

[391] Cohen J, Ni J. Semi-Supervised Learning for Anomaly Classification Using Partially Labeled Subsets. Journal of Manufacturing Science and Engineering. 2022;144(6):1-9.

[392] Serin G, Sener B, Ozbayoglu AM, Unver HO. Review of tool condition monitoring in machining and opportunities for deep learning. International Journal of Advanced Manufacturing Technology. 2020;109(3-4):953-74.

[393] Liu J, Guo F, Zhang Y, Hou B, Zhou H. Defect classification on limited labeled samples with multiscale feature fusion and semi-supervised learning. Applied Intelligence. 2021;52:8243-58. Available from: https://doi.org/10.1007/s10489-021-02917-y.

[394] Gao RX, Wang L, Helu M, Teti R. Big data analytics for smart factories of the future. CIRP Annals. 2020 jan;69(2):668-92.

[395] Lu S, Zhou P, Wang X, Liu Y, Liu F, Zhao J. Condition monitoring and fault diagnosis of motor bearings using undersampled vibration signals from a wireless sensor network. Journal of Sound and Vibration. 2018;414:81-96. Available from: https://doi.org/10.1016/j.jsv.2017.11.007.

[396] Zhang Y, Li X, Gao L, Wang L, Wen L. Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning. Journal of Manufacturing Systems. 2018 jul;48:34-50.

[397] Ganganwar V. An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering. 2012;2(4):42-7. Available from: http://www.ijetae.com/files/Volume2Issue4/IJETAE_0412_07.pdf.

[398] Nanni L, Paci M, Brahnam S, Lumini A. Comparison of different image data augmentation approaches. Journal of Imaging. 2021;7(12).

[399] Rebuffi SA, Gowal S, Calian D, Stimberg F, Wiles O, Mann T. Data Augmentation Can Improve Robustness. In: Advances in Neural Information Processing Systems. vol. 36; 2021. p. 29935-48.

[400] Dai J, Wang J, Huang W, Shi J, Zhu Z. Machinery Health Monitoring Based on Unsupervised Feature Learning via Generative Adversarial Networks. IEEE/ASME Transactions on Mechatronics. 2020;25(5):2252-63.

[401] Zeng M, Yang Y, Luo S, Cheng J. One-class classification based on the convex hull for bearing fault detection. Mechanical Systems and Signal Processing. 2016;81:274-93. Available from: http://dx.doi.org/10.1016/j.ymssp.2016.04.001.

[402] Tong Q, Lu F, Feng Z, Wan Q, An G, Cao J, et al. A Novel Method for Fault Diagnosis of Bearings with Small and Imbalanced Data Based on Generative Adversarial Networks. Applied Sciences (Switzerland). 2022;12(14).

[403] Wu Z, Jiang H, Zhao K, Li X. An adaptive deep transfer learning method for bearing fault diagnosis. Measurement. 2020;151:107227. Available from: https://doi.org/10.1016/j.measurement.2019.107227.

[404] Wang X, Shen C, Xia M, Wang D, Zhu J, Zhu Z. Multi-scale deep intra-class transfer learning for bearing fault diagnosis. Reliability Engineering and System Safety. 2020;202(May):107050. Available from: https://doi.org/10.1016/j.ress.2020.107050.

[405] Zhou G, Zhang C, Li Z, Ding K, Wang C. Knowledge-driven digital twin manufacturing cell towards intelligent manufacturing. International Journal of Production Research. 2020;58(4):1034-51. Available from: https://doi.org/10.1080/00207543.2019.1607978.

[406] Ladj A, Wang Z, Meski O, Belkadi F, Ritou M, Da Cunha C. A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective. Journal of Manufacturing Systems. 2021;58(PB):168-79. Available from: https://doi.org/10.1016/j.jmsy.2020.07.018.

[407] Umeda Y, Ota J, Shirafuji S, Kojima F, Saito M, Matsuzawa H, et al. Exercise of digital kaizen activities based on 'digital triplet' concept. Procedia Manufacturing. 2020;45:325-30. Available from: https://doi.org/10.1016/j.promfg.2020.04.025.

[408] Qiu H, Lee J, Lin J, Yu G. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal of Sound and Vibration. 2006;289(4-5):1066-90.

[409] Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-morello B, Zerhouni N, et al. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In: IEEE International Conference on Prognostics and Health Management; 2012. p. 1-9.

[410] Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mechanical Systems and Signal Processing. 2015;64-65(2015):100-31.

[411] Bechhoefer E. Condition Based Maintenance Fault Database for Testing of Diagnostic and Prognostics Algorithms. Oak Brook: MFPT; 2012. Available from: https://www.mfpt.org/fault-data-sets/.

[412] Huang H, Baddour N. Bearing vibration data collected under time-varying rotational speed conditions. Data in Brief. 2018 dec;21:1745-9. Available from: http://creativecommons.org/licenses/by/4.0/.

[413] Lessmeier C, Kimotho JK, Zimmer D, Sextro W. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification. In: Third European Conference of the Prognostics and Health Management Society 2016. Bilbao; 2016. p. 1-17.

[414] Wang B, Lei Y, Li N, Li N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Transactions on Reliability. 2020;69(1):401 411.

[415] Kim S, An D, Choi JH. Diagnostics 101: A tutorial for fault diagnostics of rolling element bearing using envelope analysis in MATLAB. Applied Sciences (Switzerland). 2020;10(20):1-23.

[416] Mauthe F, Hagmeyer S, Zeiler P. Creation of Publicly Available Data Sets for Prognostics and Diagnostics Addressing Data Scenarios Relevant to Industrial Applications. International Journal of Prognostics and Health Management. 2021;12(2):1-20.

[417] Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. 1996;12(4):5-33.

[418] Kahn BK, Strong DM, Wang RY. Information Quality Benchmarks: Product and Service Performance. Communications of the ACM. 2002;45(4):184-92.

[419] Hertzog MA. Considerations in Determining Sample Size for Pilot Studies. Research in Nursing & Health. 2008;31:180-91.

[420] Hackshaw A. Small studies: Strengths and limitations. European Respiratory Journal. 2008;32(5):1141-3.

[421] Boškoski P, Petrovičič J, Musizza B, Juriičić D. Detection of lubrication starved bearings in electrical motors by means of vibration analysis. Tribology International. 2010;43(9):1683-92.

[422] Cubillas D, Olave M, Llavori I, Ulacia I, Larrañaga J, Zurutuza A, et al. A novel formulation for radial fretting wear: Application to false brinelling in thrust bearings. Wear. 2022;488-489(June 2021):1-13.

[423] Kong Z, Tang B, Deng L, Liu W, Han Y. Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. Renewable Energy. 2020;146:760-8. Available from: `https://doi.org/10.1016/j.renene.2019.07.033`.

[424] Abedjan Z, Chu X, Deng D, Fernandez RC, Ilyas IF, Ouzzani M, et al. Detecting data errors: Where are we and what needs to be done? In: Proceedings of the VLDB Endowment; 2016. p. 993-1004.

[425] Stetco A, Dinmohammadi F, Zhao X, Robu V, Flynn D, Barnes M, et al. Machine learning methods for wind turbine condition monitoring: A review. Renewable Energy. 2019 apr;133:620-35.

[426] Malla C, Panigrahi I. Review of Condition Monitoring of Rolling Element Bearing Using Vibration Analysis and Other Techniques. Journal of Vibration Engineering and Technologies. 2019;7(4):407-14. Available from: `https://doi.org/10.1007/s42417-019-00119-y`.

[427] Hummel P, Braun M, Dabrock P. Own Data? Ethical Reflections on Data Ownership. Philosophy and Technology. 2021;34(3):545-72.

[428] Zhang Y, Wang J, Zhang X. Conciseness is better: Recurrent attention LSTM model for document-level sentiment analysis. Neurocomputing. 2021;462:101-12. Available from: `https://doi.org/10.1016/j.neucom.2021.07.072`.

[429] Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys. 2021;54(6).

[430] Chuprov S, Viksnin I, Kim I, Reznikand L, Khokhlov I. Reputation and Trust Models with Data Quality Metrics for Improving Autonomous Vehicles Traffic Security and Safety. In: 2020 Systems Security Symposium (SSS); 2020. p. 1-8.

[431] Acharya S, Mekker M. Importance of the reputation of data manager in the acceptance of connected vehicles. Communications in Transportation Research. 2022;2(January):100053. Available from: `https://doi.org/10.1016/j.commtr.2022.100053`.

[432] Chhetri SR, Rashid N, Faezi S, Faruque MAA. Security trends and advances in manufacturing systems in the era of industry 4.0. IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers, ICCAD. 2017;2017-November:1039-46.

[433] Bagheri B, Rezapoor M, Lee J. A unified data security framework for federated prognostics and health management in smart manufacturing. Manufacturing Letters. 2020;24:136-9. Available from: `https://doi.org/10.1016/j.mfglet.2020.04.011`.

[434] Herranz J, Matwin S, Nin J, Torra V. Classifying data from protected statistical datasets. Computers and Security. 2010;29(8):875-90.

[435] David M, David FR, David F. Mission statement theory and practice: a content analysis and new direction. International Journal of Business, Marketing, & Decision Science. 2014;7(1):95-110. Available from: `http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=97940355&site=eds-live`.

[436] Straker K, Wrigley C. From a Mission Statement to a Sense of Mission: Emotion Coding to Strengthen Digital Engagements. Journal of Creating Value. 2018;4(1):82-109.

[437] Liu J, Qu F, Hong X, Zhang H. A Small-Sample Wind Turbine Fault Detection Method With Synthetic Fault Data Using Generative Adversarial Nets. IEEE Transactions on Industrial Informatics. 2019 jul;15(7):3877-88.

[438] Alqudah R, Al-Mousa AA, Abu Hashyeh Y, Alzaibaq OZ. A systemic comparison between using augmented data and synthetic data as means of enhancing wafermap defect classification. Computers in Industry. 2023;145:103809. Available from: `https://doi.org/10.1016/j.compind.2022.103809`.

[439] Casaburo A, Petrone G, Franco F, De Rosa S. A Review of Similitude Methods for Structural Engineering. American Society of Mechanical Engineers (ASME); 2019.

[440] Williams LS. The mission statement: A corporate reporting tool with a past, present, and future. Journal of Business Communication. 2008;45(2):94-119.

[441] Alegre I, Berbegal-Mirabent J, Guerrero A, Mas-Machuca M. The real mission of the mission statement: A systematic review of the literature. Journal of Management and Organization. 2018;24(4):456-73.

[442] Brace W, Cheutet V. A framework to support requirements analysis in engineering design. Journal of Engineering Design. 2012;23(12):876-904.

[443] Coutinho CP, Baptista AJ, Dias Rodrigues J. Reduced scale models based on similitude theory: A review up to 2015. Engineering Structures. 2016 jul;119:81-94.

[444] AIAG & VDA. AIAG & VDA FMEA Handbook. Automotive Industry Action Group; 2019.

[445] IEC. IEC 61025 Fault Tree Analysis (FTA). Geneva: International Electrotechnical Commission; 2006.

[446] Yang W, Tavner PJ, Crabtree CJ, Wilkinson M. Cost-effective condition monitoring for wind turbines. IEEE Transactions on Industrial Electronics. 2010;57(1):263-71.

[447] García Márquez FP, Lewis RW, Tobias AM, Roberts C. Life cycle costs for railway condition monitoring. Transportation Research Part E: Logistics and Transportation Review. 2008;44:1175-87.

[448] Yang W, Tavner PJ, Crabtree CJ, Feng Y, Qiu Y. Wind turbine condition monitoring: technical and commercial challenges. Wind Energy. 2014;17:673-93.

[449] Wu Z, Liu W, Nie W. Literature review and prospect of the development and application of FMEA in manufacturing industry. International Journal of Advanced Manufacturing Technology. 2021;112(5-6):1409-36.

[450] Doggett AM. Root Cause Analysis: A Framework for Tool Selection. Quality Management Journal. 2005;12(4):34-45.

[451] Vogl GW, Weiss BA, Donmez MA. Standards Related to Prognostics and Health Management (PHM) for Manufacturing Standards Related to Prognostics and Health Management (PHM) for Manufacturing; 2014.

[452] Mulenburg GM, Gundo DP. Design by Prototype: Examples from the National Aeronautics and Space Administration. In: 11th International Conference on the Management of Technology. NASA; 2002. .

[453] Rothhaar PM, Murphy PC, Bacon BJ, Gregory IM, Grauer JA, Busan RC, et al. NASA Langley distributed propulsion VTOL tiltwing aircraft testing, modeling, simulation, control, and flight test development. In: 14th AIAA aviation technology, integration, and operations conference; 2014. p. 2999.

[454] Stone RB, Tumer IY, Van Wie M. The Function-Failure Design. Journal of Mechanical Design. 2005;127.

[455] Kurtoglu T, Tumer IY. A Graph-Based Fault Identification and Propagation Framework for Functional Design. Journal of Mechanical Design. 2008;130(May).

[456] Papakonstantinou N, Sierla S, Tumer IY, Jensen DC. Using Fault Propagation Analyses For Early Elimination of Unreliable Desing Alternatives of Complex Cyber-Physical Systems. In: Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference; 2012. p. 1-9.

[457] Irshad L, Onan Demirel H, Tumer IY. Automated generation of fault scenarios to assess potential human errors and functional failures in early design stages. Journal of Computing and Information Science in Engineering. 2020;20(5).

[458] Goldberg BE, Everhart K, Stevens R, Babbitt, N I, Clemens P, Stout L. Systems Engineering "Toolbox" for Design-Oriented Engineers. Marshall Space Flight Center: MARshall Space Flight Center; 1994.

[459] Scriboni M. FMECA and FTA analysis for industrial and collaborative robots [Master's Degree in Mechatronic Engineering]. Politecnico De Torino; 2020.

[460] Song LK, Wen J, Fei CW, Bai GC. Distributed collaborative probabilistic design of multi-failure structure with fluid-structure interaction using fuzzy neural network of regression. Mechanical Systems and Signal Processing. 2018;104:72-86. Available from: `https://doi.org/10.1016/j.ymssp.2017.09.039`.

[461] Abdelghany M, Tahar S. Cause-Consequence Diagram Reliability Analysis Using Formal Techniques with Application to Electrical Power Networks. IEEE Access. 2021;9:23929-43.

[462] Loutas TH, Sotiriades G, Kalaitzoglou I, Kostopoulos V. Condition monitoring of a single-stage gearbox with artificially induced gear cracks utilizing on-line vibration and acoustic emission measurements. Applied Acoustics. 2009;70(9):1148-59. Available from: `http://dx.doi.org/10.1016/j.apacoust.2009.04.007`.

[463] Zhao X, Zuo MJ, Liu Z, Hoseini MR. Diagnosis of artificially created surface damage levels of planet gear teeth using ordinal ranking. Measurement: Journal of the International Measurement Confederation. 2013;46(1):132-44. Available from: `http://dx.doi.org/10.1016/j.measurement.2012.05.031`.

[464] Bittencourt AC, Saarinen K, Sander-Tavallaey S. A data-driven method for monitoring systems that operate repetitively - Applications to wear monitoring in an industrial robot joint. In: 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. vol. 8; 2012. p. 198-203.

[465] Bierwagen GP, He L, Li J, Ellingson L, Tallman DE. Studies of a new accelerated evaluation method for coating corrosion resistance - thermal cycling testing. Progress in Organic Coatings. 2000;39(1):67-78.

[466] Tao L, Lu C, Yang C. Battery capacity degradation prediction using similarity recognition based on modified dynamic time warping. Structural Control and Health Monitoring. 2018;25(1):1-11.

[467] Elmenreich W. An introduction to sensor fusion. Vienna University of Technology; 2002. February. Available from: `http://www.vmars.tuwien.ac.at/documents/intern/805/elmenreich_sensorfusionintro.pdf`.

[468] Mandolla C, Petruzzelli AM, Percoco G, Urbinati A. Building a digital twin for additive manufacturing through the exploitation of blockchain: A case analysis of the aircraft industry. Computers in Industry. 2019;109:134-52. Available from: `https://doi.org/10.1016/j.compind.2019.04.011`.

[469] He B, Li T, Xiao J. Digital Twin-Driven Controller Tuning Method for Dynamics. Journal of Computing and Information Science in Engineering. 2021;21(3).

[470] Vassalos D. Physical modelling and similitude of marine structures. Ocean Engineering. 1999;26:111-23.

[471] Galilei G. Dialogues Concerning Two New Sciences; 1638.

[472] Bottasso CL, Campagnolo F, Petrović V. Wind tunnel testing of scaled wind turbine models: Beyond aerodynamics. Journal of Wind Engineering and Industrial Aerodynamics. 2014 apr;127:11-28.

[473] De A Martins R. The origin of dimensional analysis. Journal of the Franklin Institute. 1981;311(5):331-7.

[474] Rayleigh L. The principle of similitude. Nature. 1915;1(1):396-7.

[475] Zhang W, Luo Z, Li Y, Zhu Y. A novel method to determine the coupling scaling laws of vibration characteristics for rotor-bearing systems. JVC/Journal of Vibration and Control. 2020.

[476] Tukey JW. Exploratory Data Analysis. 1st ed. Reading; 1977.

[477] Stine RA, Heyse JF. Non-parametric estimates of overlap. Statistics in Medicine. 2001;20(2):215-36.

[478] SKF. Deep groove ball bearings; 2023. "[Accessed 25th April 2023]". Available from: `https://www.skf.com/us/products/rolling-bearings/ball-bearings/deep-groove-ball-bearings`.

[479] Koyo. Ball & Roller Bearings: Failures, Causes and Countermeasures. Osaka; 2021. "[Accessed 25th April 2023]". Available from: `https://koyo.jtekt.co.jp/en/support/catalog-download/uploads/catb3001ex.pdf#page=5`.

[480] Dwyer-Joyce RS. The Effects of Lubricant Contamination on Rolling Bearing Performance [Thesis]. Imperial College of Science, Technology, and Medicine; 1993.

[481] Lugt PM, Morales-Espejel GE. A review of elasto-hydrodynamic lubrication theory. Tribology Transactions. 2011;54(3):470-96.

[482] Corporation N. Calculating Grease Quantity, Frequency; 2001. "[Accessed 17th June 2022]". Available from: `https://www.machinerylubrication.com/Read/31524/calculate-grease-quantity`.

[483] Maru MMÃ, Castillo RS, Padovese LR. Study of solid contamination in ball bearings through vibration and wear analyses. Tribology International. 2007;40(1):433-40.

[484] Hariharan V, Srinivasan PSS. Condition monitoring studies on ball bearings considering solid contaminants in the lubricant. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2010 jan;224(8):1727-48.

[485] ISO. ISO 4406:2021: Hydraulic fluid power - Fluids - Method for coding the level of contamination by solid particles. ISO; 2021.

[486] Morales-Espejel GE, Gabelli A. The progression of surface rolling contact fatigue damage of rolling bearings with artificial dents. Tribology Transactions. 2015 jan;58(3):418-31.

[487] Ueda T, Mitamura N. Mechanism of dent initiated flaking and bearing life enhancement technology under contaminated lubrication condition. Part II: Effect of rolling element surface roughness on flaking resulting from dents, and life enhancement technology of rolling bearings under contaminated lubrication condition. Tribology International. 2009;42(11-12):1832-7. Available from: `http://dx.doi.org/10.1016/j.triboint.2008.12.010`.

[488] Huang NE, Shen Z, Long SR, Wu MC, Snin HH, Zheng Q, et al. The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 1998;454(1971):903-95.

[489] Zhao X, Patel TH, Zuo MJ. Multivariate EMD and full spectrum based condition monitoring for rotating machinery. Mechanical Systems and Signal Processing. 2012;27(1):712-28. Available from: `http://dx.doi.org/10.1016/j.ymssp.2011.08.001`.

[490] Wu J, Wu C, Lv Y, Deng C, Shao X. Design a degradation condition monitoring system scheme for rolling bearing using EMD and PCA. Industrial Management and Data Systems. 2017;117(4):713-28.

[491] Cooley JW, Tukey JW. An Algorithm for the Machine Calculation of Complex Fourier Series. Mathematics of Computation. 1965;19(90):297-301. Available from: `https://www.jstor.org/stable/2003354?seq=1&cid=pdf-`.

[492] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987;20(C):53-65.

[493] George D, Mallery M. SPSS for Windows Step-by-Step: A Simple Guide and Reference, 17.0 update. 10th ed. Boston: Pearson; 2010.

[494] Hair J, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7th ed. Pearson, editor. Boston; 2010.

[495] Aminu IM. Strategic Orientation, Access to Finance, Business Environment and SMEs Performance in Nigeria : Data Screening and Preliminary Analysis. European Journal of Business and Management. 2014;6(35):124-32.

[496] Li J, Hao H. A review of recent research advances on structural health monitoring in Western Australia. Structural Monitoring and Maintenance. 2016;3(1):33-49.

[497] ISO. Mechanical vibration — Measurement and evaluation of machine vibration — Part 2: Land-based gas turbines, steam turbines and generators in excess of 40 MW, with fluid-film bearings and rated speeds of 1 500 r/min, 1 800 r/min, 3 000 r/min and 3 600 r/min. ISO; 2017.

[498] Harris T, Kotzalas M. Essential Concepts of Bearing Technology. 5th ed. Boca Raton: Taylor & Francis Group; 2007.

[499] Quinn AJ, Lopes-dos Santos V, Dupret D, Nobre AC, Woolrich MW. EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python. Journal of Open Source Software. 2021;6(59):2977. Available from: https://doi.org/10.21105/joss.02977.

[500] Del E, Lemoine J, Niang O, Deléchelle E. Empirical Mode Decomposition: An Analytical Approach for Sifting Process. IEEE Signal Processing Letters. 2005;12(11):764-7.

[501] Kahlman L, Hutchings IM. Effect of particulate contamination in grease-lubricated hybrid rolling bearings. Tribology Transactions. 1999;42(4):842-50.

[502] Miettinen J, Andersson P. Acoustic emission of rolling bearings lubricated with contaminated grease. Tribology International. 2000;33(11):777-87.

[503] Sahu PK, Rai RN, Kumar TCA. Grease Contamination Detection in the Rolling Element Bearing Using Deep Learning Technique. International Journal of Mechanical Engineering and Robotics Research. 2022;11(4):275-80.

[504] Yazdi AA, Rezaeepazhand J. Structural similitude for flutter of delaminated composite beam-plates. Composite Structures. 2011;93(7):1918-22.

[505] Buckingham E. On physically similar systems; illustrations of the use of dimensional equations. Physical Review. 1914;4(4):345-76.

[506] Jamadar IM, Vakharia DP. A novel approach integrating dimensional analysis and neural networks for the detection of localized faults in roller bearings. Measurement. 2016 dec;94:177-85.

[507] Jamadar IM, Vakharia DP. A Numerical Model for the Identification of the Structural Damages in Rolling Contact Bearings Using Matrix Method of Dimensional Analysis. Journal of Tribology. 2016 apr;138(2).

[508] Kumbhar SG, Edwin Sudhagar P. Fault Diagnostics of Roller Bearings Using Dimension Theory. Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems. 2021;4(1):1-10.

[509] Li L, Luo Z, He F, Ding Z, Sun K. A partial similitude method for vibration responses of rotor systems: Numerical and experimental verification. International Journal of Mechanical Sciences. 2021;208(June):106696. Available from: https://doi.org/10.1016/j.ijmecsci.2021.106696.

[510] Luo Z, Li L, He F, Yan X. Partial similitude for dynamic characteristics of rotor systems considering gravitational acceleration. Mechanism and Machine Theory. 2021;156:104142. Available from: https://doi.org/10.1016/j.mechmachtheory.2020.104142.

[511] Singhatanadgid P, Na Songkhla A. An experimental investigation into the use of scaling laws for predicting vibration responses of rectangular thin plates. Journal of Sound and Vibration. 2008;311(1-2):314-27.

[512] Rafsanjani A, Abbasion S, Farshidianfar A, Moeenfard H. Nonlinear dynamic modeling of surface defects in rolling element bearing systems. Journal of Sound and Vibration. 2009;319:1150-74.

[513] Patel VN, Tandon N, Pandey RK. A Dynamic Model for Vibration Studies of Deep Groove Ball Bearings Considering Single and Multiple Defects in Races. Journal of Tribology. 2010 10;132(4). 041101. Available from: `https://doi.org/10.1115/1.4002333`.

[514] Nakhaeinejad M, Bryant MD. Dynamic Modeling of Rolling Element Bearings With Surface. Journal of Tribology. 2011;133(January):1-12.

[515] Brewe DE, Hamrock BJ. Simplified Solution for Elliptical-Contact Deformation Between Two Elastic Solids. Journal of Lubrication Technology. 1977;99(4):485-7. Available from: `https://doi.org/10.1115/1.3453245`.

[516] Mishra C, Samantaray AK, Chakraborty G. Ball bearing defect models: A study of simulated and experimental fault signatures. Journal of Sound and Vibration. 2017;400(July):86-112.

[517] Cao H, Shi F, Li Y, Li B, Chen X. Vibration and stability analysis of rotor-bearing-pedestal system due to clearance fit. Mechanical Systems and Signal Processing. 2019;133:106275. Available from: `https://doi.org/10.1016/j.ymssp.2019.106275`.

[518] Crandall SH. The role of damping in vibration theory. Journal of Sound and Vibration. 1970;11(1):3-18.

[519] Genta G. On a Persistent Misunderstanding of the Role of Hysteretic Damping in Rotordynamics. Journal of Vibration and Acoustics. 2004;126(July):459-61.

[520] Jung SS, Lee YB, Jeon BS, Shin SH. Measurement of the loss factor and the Young's modulus in structural steel by using a laser beam reflection method. Journal of the Korean Physical Society. 2014;65(7):1024-7.

[521] ISO. ISO 5753-1:2009 Rolling bearings - Internal Clearance - Part 1: Radial internal clearance for radial bearings. Geneva: International Organization for Standardization; 2009.

[522] Chandrashekar G, Raj W, Godwin C, Paul PS. Study on the Influence of Shaft Material on Vibration in Rotating Machinery. Materials Today: Proceedings. 2018;5(5):12071-6. Available from: `https://doi.org/10.1016/j.matpr.2018.02.182`.

[523] Karna SK, Sahai R. An Overview on Taguchi Method. International Journal of Engineering and Mathematical Sciences. 2012;1:11-8. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/19879888`.

[524] Boll SF. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1979;27(2):113-20.

[525] Al-badour F, Sunar M, Cheded L. Vibration analysis of rotating machinery using time – frequency analysis and wavelet techniques. Mechanical Systems and Signal Processing. 2011;25(6):2083-101. Available from: `http://dx.doi.org/10.1016/j.ymssp.2011.01.017`.

[526] Yan R, Gao RX, Chen X. Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing. 2014;96(PART A):1-15. Available from: `http://dx.doi.org/10.1016/j.sigpro.2013.04.015`.

[527] Zhang Y, Zhao Y. Real and imaginary modulation spectral subtraction for speech enhancement. Speech Communication. 2013;55(4):509-22. Available from: `http://dx.doi.org/10.1016/j.specom.2012.09.005`.

[528] Konar P, Chattopadhyay P. Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). Applied Soft Computing Journal. 2011;11(6):4203-11. Available from: http://dx.doi.org/10.1016/j.asoc.2011.03.014.

[529] Cerrada M, Zurita G, Cabrera D, Sánchez RV, Artés M, Li C. Fault diagnosis in spur gears based on genetic algorithm and random forest. Mechanical Systems and Signal Processing. 2016;70-71:87-103. Available from: http://dx.doi.org/10.1016/j.ymssp.2015.08.030.

[530] Liu TI, Mengel JM. Intelligent monitoring of ball bearing conditions. Mechanical Systems and Signal Processing. 1992;6(5):419-31.

[531] Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement: Journal of the International Measurement Confederation. 2016;93:490-502. Available from: http://dx.doi.org/10.1016/j.measurement.2016.07.054.

[532] Li B, Chow MY, Tipsuwan Y, Hung JC. Neural-network-based motor rolling bearing fault diagnosis. IEEE Transactions on Industrial Electronics. 2000;47(5):1060-9.

[533] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.

[534] Bebis G, Georgiopoulos M. Feed-forward neural networks Why nwtwork size is so important. IEEE Potentials. 1994;13(4):27-31.

[535] Erb RJ. Introduction to Backpropagation Neural Network Computation. Pharmaceutical Research. 1993;10(2):165-70. Available from: https://doi.org/10.1023/A:1018966222807.

[536] Schmidhuber J. Deep Learning in neural networks: An overview. Neural Networks. 2015;61:85-117. Available from: http://dx.doi.org/10.1016/j.neunet.2014.09.003.

[537] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative Adversarial Networks: An Overview. IEEE Signal Processing Magazine. 2018;35(1):53-65.

[538] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997 nov;9(8):1735-80.

[539] Lu H, Barzegar V, Nemani VP, Hu C, Laflamme S, Zimmerman AT. GAN-LSTM Predictor for Failure Prognostics of Rolling Element Bearings. In: 2021 IEEE International Conference on Prognostics and Health Management, ICPHM 2021. IEEE; 2021. p. 1-8.

[540] Chen H, Liu H, Chu X, Liu Q, Xue D. Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network. Renewable Energy. 2021;172:829-40. Available from: https://doi.org/10.1016/j.renene.2021.03.078.

[541] Chen X, Zhang B, Gao D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. Journal of Intelligent Manufacturing. 2021;32(4):971-87. Available from: https://doi.org/10.1007/s10845-020-01600-2.

[542] Sonnenberg FA, Beck JR. Markov Models in Medical Decision Making: A Practical Guide. Medical Decision Making. 1993;13(4):322-38.

[543] Zhang Y, Kim CW, Tee KF. Maintenance management of offshore structures using Markov process model with random transition probabilities. Structure and Infrastructure Engineering. 2017;13(8):1068-80. Available from: http://dx.doi.org/10.1080/15732479.2016.1236393.

[544] Li Z, Fang H, Huang M, Wei Y, Zhang L. Data-driven bearing fault identification using improved hidden Markov model and self-organizing map. Computers and Industrial Engineering. 2018;116(February 2017):37-46. Available from: `https://doi.org/10.1016/j.cie.2017.12.002`.

[545] Liu J, Hu Y, Wu B, Wang Y, Xie F. A Hybrid Generalized Hidden Markov Model-Based Condition Monitoring Approach for Rolling Bearings. Sensors (Basel, Switzerland). 2017;17(5):1-19.

[546] Yuwono M, Qin Y, Zhou J, Guo Y, Celler BG, Su SW. Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model. Engineering Applications of Artificial Intelligence. 2016;47:88-100. Available from: `http://dx.doi.org/10.1016/j.engappai.2015.03.007`.

[547] Lipton ZC, Elkan C, Naryanaswamy B. Optimal thresholding of classifiers to maximize F1 measure. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014;8725 LNAI(PART 2):225-39.

[548] Louppe G. Understanding Random Forests: From Theory to Practice [PhD dissertation]. Univrsity of Liege; 2014. Available from: `http://arxiv.org/abs/1407.7502`.

[549] Kong W, Li H. Remaining useful life prediction of rolling bearing under limited data based on adaptive time-series feature window and multi-step ahead strategy. Applied Soft Computing. 2022;129:109630. Available from: `https://doi.org/10.1016/j.asoc.2022.109630`.

[550] Taylor TC. The Costs of Training New Employees, Including Hidden Expenses; 2017.

[551] Li Y. How to Estimate the Time and Cost to Train a Machine Learning Model; 2022. Available from: `https://towardsdatascience.com/how-to-estimate-the-time-and-cost-to-train-a-machine-learning-model-eb6c8d433ff7`.

[552] Wescoat E, Krugh M, Henderson A, Goodnough J, Mears L. Vibration analysis utilizing unsupervised learning. In: Procedia Manufacturing. vol. 34; 2019. p. 876-84.