UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Mecânica

LAÍS CARRER SILVA

Fault identification in ball bearings: a simulation dataset approach and signal encoder feature extraction

Identificação de falhas em rolamentos de esferas: uma abordagem por conjunto de dados simulados e por codificador de sinal para extração de atributos

CAMPINAS
2024

LAÍS CARRER SILVA

Fault identification in ball bearings: a simulation dataset approach and signal encoder feature extraction

Identificação de falhas em rolamentos de esferas: uma abordagem por conjunto de dados simulados e por codificador de sinal para extração de atributos

Thesis presented to the School of Mechanical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Mechanical Engineering in the area of Mechanics of Solids and Mechanical Design

Tese apresentada à Faculdade de Engenharia Mecânica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutora em Engenharia Mecânica, na Área Mecânica dos Sólidos e Projeto Mecânico.

Orientadora: Prof[a]. Dr[a]. Katia Lucchesi Cavalca Dedini
Coorientador: Prof. Dr. Tiago Henrique Machado

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA LAÍS CARRER SILVA E ORIENTADA PELA PROFA. DRA KATIA LUCCHESI CAVALCA DEDINI.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca da Área de Engenharia e Arquitetura
Elizangela Aparecida dos Santos Souza - CRB 8/8098

Si38f

Silva, Laís Carrer, 1988-
Fault identification in ball bearings : a simulation dataset approach and signal encoder feature extraction / Laís Carrer Silva. – Campinas, SP : [s.n.], 2024.

Orientador: Katia Lucchesi Cavalca Dedini.
Coorientador: Tiago Henrique Machado.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Faculdade de Engenharia Mecânica.

1. Rolamento de esferas. 2. Aprendizado de máquinas. 3. Lubrificação elastohidrodinâmica. 4. Extração de características (Inteligência artificial). 5. Análise de falha de equipamento. I. Dedini, Katia Lucchesi Cavalca, 1963-. II. Machado, Tiago Henrique, 1986-. III. Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Mecânica. IV. Título.

Informações Complementares

Título em outro idioma: Identificação de falhas em rolamentos de esferas : uma abordagem por conjunto de dados simulados e por codificador de sinal para extração de atributos
Palavras-chave em inglês:
Ball-bearings
Machine learning
Elastohydrodynamic lubrication
Feature extraction (Artificial intelligence)
Equipment failure analysis
Área de concentração: Mecânica dos Sólidos e Projeto Mecânico
Titulação: Doutora em Engenharia Mecânica
Banca examinadora:
Katia Lucchesi Cavalca Dedini [Orientador]
Mateus Giesbrecht
Rafael de Oliveira Teloli
Gregory Bregion Daniel
João Paulo Dias
Data de defesa: 16-07-2024
Programa de Pós-Graduação: Engenharia Mecânica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-0015-837X
- Currículo Lattes do autor: http://lattes.cnpq.br/4653315203591007

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA

TESE DE DOUTORADO ACADÊMICO

Fault identification in ball bearings: a simulation dataset approach and signal encoder feature extraction

Identificação de falhas em rolamentos de esferas: uma abordagem por conjunto de dados simulados e por codificador de sinal para extração de atributos

Autora: Laís Carrer Silva
Orientadora: Profª. Drª. Katia Lucchesi Cavalca Dedini
Coorientador: Prof. Dr. Tiago Henrique Machado

A Banca Examinadora composta pelos membros abaixo aprovou esta Tese:

Profª. Drª. Katia Lucchesi Cavalca Dedini, Presidente
DSI /FEM/Unicamp

Prof. Dr. Gregory Bregion Daniel
DSI /FEM/Unicamp

Prof. Dr. João Paulo Dias
The School of Engineering/Shippensburg University

Prof. Dr. Mateus Giesbrecht
DEEB /FEEC/Unicamp

Prof. Dr. Rafael de Oliveira Teloli
SUPMICROTECH-ENSMM

A Ata de Defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 16 de julho de 2024.

Dedication

Para os meus pais, Jurandir e Maria Luísa.

Acknowledgements

I would like to express my gratitude to my advisor, Prof. Katia Cavalca, for her support and expertise in the development of this research. In addition, for her patience and understanding throughout the entire process.

Thanks to my co-advisor, Prof. Tiago Machado, for his assistance with the structure of this work and for providing the meaningful insights necessary to complete this research.

I would like to acknowledge the support of my colleagues from LAMAR for the technical discussions which contributed directly and indirectly to this work.

My partner, Gustavo, you helped me so much, with so many things, and really pushed me forward when I needed it most.

To my parents, Jurandir and Maria Luisa, who were always by my side and never doubted me. You two set an example of how to be courageous, so I could not back out.

I also would like to thank Universidade Estadual de Campinas - UNICAMP and Faculdade de Engenharia Mecânica – FEM for providing the facilities and means for this research.

This study was financed by Petróleo Brasileiro S.A.- PETROBRAS.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

Abstract

In recent years, machine learning applications have gained much attention in numerous fields. In the industry, it opened a wide range of possibilities to monitor, identify, and predict components health status. Rolling element bearings are present in a wide range of applications and to predict their conditions it is necessary to prevent catastrophic failures and unnecessary production interruption. To tackle this, algorithms and computational capability had to improve, and the critical requirement is good quality and proper quantity of fault information. In this context, the present work proposes to create a rolling element bearing fault data set, with the simulated signal. To build reliable computational models, the rolling bearing is modeled by its reaction force, and its parameters represent the elastohydrodynamic contact between ball elements and internal races. The fault is represented as an imposed displacement and velocity in the established locations, and a smooth function is proposed to include varied faults in the model. The simulated vibration envelope is obtained by the Hilbert transform and its spectrum composes the training data set. To extract necessary information from the envelope spectrum and make this comparable to other measured signals, it was necessary to evaluate the spectrum by pieces and guarantee the reproducibility of the procedure in all samples. A pipeline with an encoder that extracts features from the sample envelope spectrum and applies it to build a machine learning model is proposed. A support vector machine algorithm is the classifier employed due to its advantages in high feature dimension performance and good generalization capability. The elastohydrodynamic contact ball bearings model with the proposed fault functions are compared with experiments in the literature. The machine learning models are tested with the Paderborn University accelerated life damage bearings dataset, and the results show a compatible prediction compared with real measured fault signal, and the potential to be applied in fault classification in further applications.

Key Word: Ball-bearings, Machine learning, Elastohydrodynamic lubrication, Feature extraction (Artificial intelligence), Equipment Failure Analysis.

Resumo

A aplicação de aprendizado de máquina ganhou muita atenção nos últimos anos, em diversos campos. Na indústria, abriu uma ampla gama de possibilidades para monitorar, identificar e prever a integridade dos componentes. Rolamentos de elementos rolantes são utilizados em uma ampla variedade de aplicações, e prever suas condições é necessário para prevenir falhas catastróficas e interrupções desnecessárias na produção. Para enfrentar esse desafio, os algoritmos e a capacidade computacional tiveram que melhorar, e agora o requisito crítico é a boa qualidade e quantidade adequada de informações sobre falhas reais. Nesse contexto, o presente trabalho propõe criar um conjunto de dados de falhas em rolamentos esferas com sinais simulados. Para construir modelos computacionais confiáveis, o rolamento é modelado por sua força de reação, e seus parâmetros representam o contato elastohidrodinâmico entre os elementos esféricos e as pistas internas. A falha é representada como um deslocamento e uma velocidade impostos nas localizações determinadas, e uma função suave é proposta para incluir várias falhas no modelo. O envelope do sinal de vibração simulada é obtido pela transformada de Hilbert e seu espectro compõe o conjunto de dados de treinamento. Para extrair as informações necessárias do espectro do envelope e torná-las comparáveis a outros sinais medidos, foi necessário avaliar o espectro por partes e garantir a reprodutibilidade do procedimento em todas as amostras. É proposto uma linha de processos com um codificador que extrai características do espectro do envelope das amostras e as aplica para construir um modelo de aprendizado de máquina. Um algoritmo de máquina de vetores de suporte é o classificador empregado devido às suas vantagens em desempenho em alta dimensão de características e boa capacidade de generalização. O modelo de rolamentos de contato elastohidrodinâmico com as funções de falha propostas é comparado com experimentos na literatura. Os modelos de aprendizado de máquina são testados com o conjunto de dados de rolamentos com falhas de vida acelerada da Universidade de Paderborn, e os resultados mostram uma previsão compatível com o sinal de falha real medido, com potencial para ser aplicado na classificação de falhas em futuras aplicações.

Palavras-chave: Rolamento de esferas, Aprendizado de máquinas, Lubrificação elastohidrodinâmica, Extração de características (Inteligência artificial), Análise de Falha de Equipamento.

List of Figures

List of Tables

List of Symbols and Abbreviations

b	Support vectors bias
C	Penalty hyperparameter
C_d	Fault or defect maximum depth
C_r	Radial clearance
d	Exponent
d_e	Spherical element diameter
d_i	Boundary side of the hyperplane
d_p	Pitch diameter
D	Damping parameter
D_h	Housing node damping
D_s	Shaft damping
D_{bl}	Linearized bearing damping
f	Frequency in Hz
f_b	Bearing reaction force
f_{ord}	Frequency order
$\mathrm{F_R}$	Force in radial direction
F_{me}	Unbalanced force
h	Fluid film thickness
K	Kernel function
K	Equivalent parameter for stiffness
K_h	Housing node stiffness
K_H	Dry contact stiffness
K_s	Shaft Stiffness
K_{bl}	Linearized bearing stiffness
m_e	the residual mass
m_h	Internal housing mass
m_s	Shaft mass
M_{gj}	Gyroscopic moment of the j^{th} Element
N	Number of features
p	Contact pressure

$Q_{i,o}$	Load in the inner and outer contact
r	Polynomial function independent coefficient
r_e	Element radius
r_i	Inner race radius
r_o	Outer race radius
S	Feature
t	Tempo
u_m	Surfaces relative velocity.
w	Weight vector
x_h	Vertical housing position
x_s	Vertical shaft position
X	Signal amplitude
y_h	Horizontal housing position
y_s	Horizontal shaft position
Z	Number of rolling elements

α	Angular contact angle
β	Fault operator
γ	Kernel wideness of the function
Γ	Summation of v in the load zone
$\delta_{i,o}$	Displacement in the inner and outer contact
δ_{max}	Displacement in the element with maximum load.
δ_j	Displacement in the j^{th} Element
$\Delta\theta_d$	Fault angular span
Δ_{fnh}	Frequency order span
ΔF	Offset force
Δs	Fault extension
ε	Eccentric mass radius
ϵ	Vector of ones
θ_{d0}	Fault initial angular position
θ_{df}	Fault final angular position

θ_s	Spin angle
λ	Slip coefficients
μ	Lubricant viscosity
ξ_i	Outlier control variable
ρ	Density
ϱ	Binary variable
σ	Dual coefficients
φ	Azimuth position
ϕ	Mapping function in feature space
ω_c	Cage rotational speed
ω_{spin}	Element spin rotation
Ω	System rotational speed
$\mathscr{f}$	Feature central frequency order
$\mathscr{f}_{cut}$	Frequency order cut off
$\mathscr{o}$	Features overlap

AI	Artificial Intelligence
DOF	Degree of Freedom
EHD	Elastohydrodynamic
EMD	Empirical Mode Decomposition
ES	Envelope Spectrum
FFT	Fast Fourie Transform
FIR	Finite Impulse Response
FK	Fast computation of Kurtogram
HFRT	High Frequency Resonance Technique
HT	Hilbert Transform
IMF	Intrinsic Mode Frequency
IoT	Internet of Things
IRF	Inner Race Fault
MBS	Multi-Body dynamics Simulation
ML	Machine Learning
MLMI	Multi-Level Multi-Integration

MLP	Multi-Layer Perceptron
ORDWT	Over-Complete Rational Dilation Wavelet Transform
ORF	Outer Race Fault
PSD	Power Spectral Density
RE	Rolling Element
REB	Rolling Element Bearings
RPM	Rotation per Minute
SK	Spectral Kurtosis
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TEO-AK	Temporal Energy Operated Auto-correlated Kurtosis
VMD	Variational Mode Decomposition
WPT	Wavelet Packet Transform

Summary

1 INTRODUCTION

Ball bearings are present in most rotating machine applications and other countless devices ranging from common domestic appliances to simple tools, skateboards, and even toys. Furthermore, these components are extensively used in industrial machinery including trains, torque converters, and wind turbines. To ensure optimal performance across diverse applications and operating conditions, scientists, engineers, and designers employ theoretical models, to enhance performance, predict reliability and robustness under various operating conditions and optimize resources.

The performance of these components is crucial for efficient operations as well as establishing effective maintenance practices. Both are crucial to prevent premature crashes and catastrophic failures. The monitoring tools are constantly improved aiming for better applicability in health monitoring of complex systems and promising applications, wind turbines having been the recent focus of new advances. It has been reported that bearing fault is the primary cause of downtime in gearboxes and generators (Azevedo et al, 2016). The mechanisms and conditions for bearing faults are commonly separated into two categories, localized/incipient or distributed defects.

Localized faults are usually referred to as pits, cracks, or spalls. The spalling resulting from fatigue is the most common cause of premature bearing failures (GUPTA, PRADHAN, 2017; ZHANG et al., 2022). Fatigue failure leads to the material removal from the inner race, the outer race, or the rolling elements. The types of fatigue are commonly divided into three groups, surface distress, pitting, and spalling (FREITAG, 2014).

The distributed fault, usually represented as waviness and surface roughness, can be from initial fault propagation but typically it is due to manufacturing errors (SUNNERSJÖ, 1985). Different types of failure and its induced factors can overlap, or one type may start, and the new condition (with the fault) then leads to another failure mode (HOWARD, 1994).

Recent industrial production and quality assessment prevents and mitigates manufacturing inadequacies. Therefore, in this work, the representation of fault in the ball bearing is a localized spall. The objective is to develop a fault model and implement diagnostic tools capable of identifying potential faults at their inception. The aim is to quickly detect anomalies and alert the user. The maintenance schedule can then be programmed accordingly, while simultaneously monitoring the fault behavior to anticipate further failures.

Both early stop and catastrophic failures are a hazard to production. Early stops can result in undesired delays and costs, while catastrophic failures lead to an unscheduled stop, compromising deadlines, and budgets, and risk damaging other components. In this context, the Internet of Things (IoT) plays a major role in the 4.0 industry. Its application in predictive maintenance aims to reduce unplanned downtime by identifying potential equipment failure (SOORI; AREZOO; DASTRES, 2023). The correct use of new techniques, constantly referred to as the 4.0 industry, allows real-time analysis for predictive maintenance, integrating health monitoring into the maintenance of equipment, avoiding failures, reducing downtime, and possibly extending components' lifespan.

For this purpose, is essential to develop and maintain reliable fault models and tools to evaluate and classify the equipment components conditions. One possible way to conduct these monitoring procedures is by applying Machine Learning (ML) models trained for a specific component and under its possible operational conditions. The processes usually consist of, but it is not restricted to, the following steps:

1. Selecting appropriate sensors and positioning: for ball bearing fault identification accelerometers are usually placed in the housing, in one or two directions. It is important to guarantee that the information has desirable properties, with a suitable resolution, sufficient amount of data, and proper filtration to avoid aliasing and leakage on the digital representation of the signal.
2. Data pre-processing usually includes other filtering techniques, demodulations, and scaling of the data. It is necessary to be aware of the risks of removing important information about the conditions together with undesirable noise.
3. Extracting and selecting features from the data is a crucial stage. The data is summarized and encoded to represent normal and abnormal operations. Most authors usually apply each sample signal as a distribution and synthesize this data into statistical features (such as root mean squared, variance, kurtosis, and other profile indicators, including shape and crest factor).

4. Develop the machine learning model: The ML will be properly trained with the selected features. This involves selecting the better suitable algorithm for the task and setting its hyperparameters, if applicable. When needed, the definition of the hyperparameters can result in a time-consuming and usually computation-costly optimization or search tool. There are a few good practices to prevent overfitting the model, and to guarantee its generalization capability for the variety of input conditions expected to be evaluated, when in use.
5. After tests and identification of the model accuracy (or other metrics) rate, the model can be applied to unlabeled data to, within certain confidence, diagnose Rolling Element Bearings (REB) health status.

Pre-processing data is fundamental to understanding its behavior in the desirable monitoring conditions. As widely known, the fault in REB modulates the vibration time signals and can excite high frequencies in the initial stage. Demodulating the signal to identify those frequencies is one of the most used paths for fault identification. Since its development, in the 1970s decade, the High Frequency Resonance Technique (HFRT) has been largely used, what today is usually referred to as envelope analysis. Other methods for improving and filtering the signal, such as kurtogram, are applied to extract the most informative part, without compromising important information.

The main obstacle for ML applications is the deficiency of good quality signals for a variety of conditions, as mentioned in (RANDALL ; ANTONI, 2011). Another recurrent issue is the fact that these methods heavily rely only on statistical patterns, in addition, usually, the model is specific for one type of machine and its configuration, resulting in a model not being robust enough for different sets of inputs. To tackle this issue, this work proposes a non-dimensional feature extraction that can be representative of a broader machine configuration.

One of the major setbacks to exploring machine-learning algorithms for classification is the need for good-quality data from faulty bearings. It is challenging to acquire damaged bearing signals since the majority of data is from health operational conditions. Even when fault signals are obtained the variety of conditions is limited. As a solution, it is proposed a fault model for lubricated ball bearings. The fault is a spall located in one of the regions, Outer Race Fault (ORF), Inner Race Fault (IRF), or rolling element fault.

This model is applicable for a variety of systems and implementations, by simply including the effect of the bearing as a reaction force in the center of the bearing node. The reduced model comprehends Elastohydrodynamic (EHD) lubrication (NONATO, CAVALCA, 2014) and it is especially advantageous for creating datasets based on the machine configuration and the bearing operational condition.

The process of analyzing the data, both experimental and simulated, implies training ML models to perform the diagnoses and identification in a given system. Among a vast number of theories and algorithms for machine learning models, this thesis performs a brief analysis of constantly used methods found in the literature and motivations for selecting a Support Vector Machine (SVM) as the model classifier.

The proposed bearing model can easily be applied to simulate the fault under different sets of conditions and this data is incorporated with experimental health and fault conditions, to obtain a comprehensive dataset. The features are extracted from the Envelope Spectrum (ES), where the amplitudes of the spectrum are normalized and the frequencies are adimensionalized by the rotating frequency. Then the ES is encoded. The encoding process is a feature selection that designates the scaled amplitude per range of frequencies. The parameters for this feature selection are defined by Bayesian optimization, simultaneously with the classifier hyperparameters tuning.

The thesis is organized into five chapters, including this introduction, and subdivided into specific topics as follows:

- 2. Literature Review: This chapter revisits the main foundation and latest works on rolling elements bearings modeling, fault modeling, signal analyses for fault identification, and feature extraction for diagnoses based on machine learning models.
- 3. Methodology: Representation of the EHD lubricated contact model to obtain the reduced force applied in the simulated ball bearings time response. The fault models and their expected behavior with or without the defects and faults. The theory of Hilbert Transform and the classification algorithm, SVM, are briefly presented. The encoding method for feature selection is proposed. The methodology for evaluating the machine learning model is divided into three: first applying the same nested cross-validation as literature, with the same data set, to evaluate the applicability of the proposed encoding method for feature

extraction, and selection. The second one is training the ML model with simulated faults and testing with the measured real fault signals. The last one is to evaluate if augmenting a real data set with simulated fault samples can improve the accuracy.

- 4. Description of Experimental and Simulated Data: This section presents the literature used to validate the model, the experimental public dataset (from Paderborn University), and the configuration for the simulated data generator used to validate the method for feature extraction and its capability of model training.
- 5. Results and Discussion: This section shows the model validation with the literature and experimental data. The process of building the ML model with the steps of feature analysis and extraction, learning curve, parameters, and feature optimization, leads to the final model presentation.
- 6. Conclusion: The synthesis of the foremost findings and contributions of this thesis are presented in the conclusion, showing the advantages of the proposed methodology, next steps, and recommendations.

The contributions of this work are the bearing model with a smooth fault path and lubricated contact, applying the elastohydrodynamic reduced model function. This model application in building a data set, with simulated data, for training a machine learning model to identify fault measurements of bearing with real faults. For this, an encoding algorithm performs a feature extraction and selection, suitable for adimensionalized data, and specifically convenient when operating with data from different sources, simulated and measured.

2 LITERATURE REVIEW

Ball bearings, as general rolling element bearings, are present in a wide range of applications, from household appliances to heavy machinery. The design and construction of ball bearings have continued to evolve and improve over time, with advancements in computer modeling and simulation to optimize their performance. The expertise in ball bearings' behavior and their interactions with other components is necessary for reliable diagnoses.

Measured signals must be pre-processed, and properly interpreted. Therefore, various tools have been proposed, based on mathematical, statistical, or empirical methods. The goal is to de-noise, de-modulate, or enhance the signature of failures and incipient defects. The technological advance for maintenance based on predictive models and conditioning monitoring expertise have driven the status up to nowadays research, namely, a widespread application of automatized identification, and Machine Learning (ML) algorithms.

Machine learning applications do not surpass past identification techniques. Rather, those techniques are the pathway for constructing models that are both general and precise for classification tasks. These signal analyses and theoretical bearing models leverage information to extract and select the most suitable features for the task. Feature extraction, a crucial step in ML, is sometimes overlooked, with the focus predominantly on the algorithms themselves.

In the present work, the feature extraction is obtained from bearing signal analysis, from both measured and simulated data. The numerical models must be reliable in representing the machinery. The theory is ground-based in contact interaction, as forces and deformation, and how to translate it to parameters and function.

2.1 Rolling bearings

One of the leading contributions in elastic deformation Hertz (1896)models the conditions when two rigid spherical surfaces are in contact. This sets a new ground for the following studies of Rolling Element Bearings (REB). The contact between the rolling element (RE) and the raceways is a critical aspect that determines its performance, durability, and efficiency. It provides a fundamental basis for modeling REBs. (HARRIS, 1991)

Lundberg and Palmgren (1949) established a connection between the applied loads and the bearing durability. The Lundberg-Palmgren theory helped increase REB material quality and lubrication contribution in bearings life after the development of the elastohydrodynamic (EHD) theory (HOWARD, 1994). Taking the research further, Jones (1960) presented a more comprehensive approach, proposing an analytical model that uses both forces and elastic deformations experienced by the internal components of the bearing.

Perret (1950) and Meldau (1951) investigate cyclic movements in bearings. Perret's work focused on symmetric arrangements of rolling elements, while Meldau provided a more comprehensive description, deriving shaft loci for ball and roller bearings in a plane perpendicular to the rotation axis. Later, Tamura and Taniguchi (1960 and 1961) conducted measurements aligned with Meldau's proposition.

Dowson (1962) suggests that under certain conditions, the contact in the bearing surfaces generating elastic deformation could lead to a significant increase in the lubricant film thickness, which in turn could reduce friction and wear, paving the way for of EHD lubrication theory. Later, a book (DOWSON; HIGGINSON, 1977) expanded this concept by presenting a comprehensive analysis of the physics and mechanics of EHD lubrication. They developed mathematical models for calculating the pressure distribution and film thickness in lubricated contacts, considering the effects of surface roughness, viscosity, and relative velocity.

Harris and Mindel (1973) applied non-linear stiffness coefficients to describe radial and axial contact forces, known as the Hertzian contact relationship. Harris (1991) focused his book on the analysis and design of rolling bearings. It provides comprehensive coverage of bearing types, their operation principles, and various factors affecting their performance. The equations presented in his book are the basis for estimating the initial forces of the bearing model developed in the present work.

(GUPTA, P. K., 1979a, b, c, d, 1975) endeavored to obtain a complete dynamic model for REB, reported by solving a generalized differential equation of motion for angular contact ball bearing. His extensive work comprehends interaction in roller-race, roller-cage and cage-raceway, lubricant drag and churning, roller skew, cage instabilities, varying the components material properties, as well as a wide range of operating conditions, adding a considerable computational complexity for that time and major difficulties for experimental verification.

Gupta’s book (1984)also comprehends geometrical unconformities, some of these conditions validated through experiments(GUPTA, P. K.; DILL; BANDOW, 1985).

In 1978, Sunnersjö discussed the main cause of noise and unsteady running in rolling bearings. His experiments confirmed the occurrence of predicted phenomena associated with varying compliance vibrations, although quantitative comparisons between theoretical and experimental results were challenging, given the influence of bearing clearance besides other components. In a similar line, Fukata et al.(1985), compared simulations and experiments, analyzing the system with two degrees of freedom and finding its sub- and super-harmonic resonances.

Kraus et al. (1987) conducted an experimental study to characterize the physical properties of lubricated roller bearings to approximate the values of bearing stiffness and damping, employing modal parameters to obtain these approximations. Following a similar approach, Lim and Singh (1990a, b, 1991, 1992), in their multiple works, proposed a stiffness matrix connecting the shaft and bearings by taking the derivative of the forces and moments for each degree of freedom, assuming dry contact.

By the 1980-decade, profuse scientific investigation had been reported on vibration signals from a regular operating ball bearing. In 1982, Igarashi and Hamada conducted experiments with artificially damaged ball bearings to analyze the frequencies that rose from the fault to compare with the natural undamaged signals from health bearings, using the Fast Fourier Transform (FFT) of the signals. It is observed the presence of several components in the spectrum. The authors even suggested that the modulation of the vibration pulse train is affected by the axial length of the defect. In Johnson’s book (1985)the focus was on the fundamental theory and mathematical modeling of contact between elastic bodies. It extensively discussed Hertzian contact theory, including the Hertzian theory for spherical and cylindrical contacts.

Lubrecht et al., (1986) show the solution of the EHD lubrication of the linear contact by the Multilevel method, reducing time and computational consumption. Venner (1991) proposed a simplification of the Lubrecht system. Later Verner and Lubrecht (2000) came out with the method of Multi-Level Multi-Integration (MLMI) for the elastic deformation equation of the contact. The EHD theory aims to accurately represent the lubricated contact under the conditions of the full fluid film, where both surfaces are completely separated under load and relative motion between them.

Wijnant (1998) studied the ellipticity of the contact area and starvation situations in lubrication, presenting an approximation of stiffness and damping for the lubricated contact. Subsequently, this methodology is applied to a complete bearing, Wijnant et al. (1999), proposing an approximation of the lubricated contact by a mass-spring-damper model.

The EHD lubrication model involves the solution of the equations of motion, iteratively with Reynolds equation and Hertz contact equation at each step of time, leading to a complex and time-consuming numerical simulation. To overcome this issue, Nonato and Cavalca (2014 and 2010) proposed a reduced nonlinear force model for the lubricated contact of a bearing, as a function of displacement and residual force due to oil film'.

Sequentially, Carvalho (2010)analyzed the influence of different inner race clearance and rotation speed values on the load distribution in a radial ball bearing. Based on (HARRIS, 1991), Radaelli (2013) modeled an angular contact ball bearing with five degrees of freedom under radial and thrust loads, including inertial forces and gyroscopic moments.

Bizarre et al.(2016) integrate the work of Radaelli (2013) and Nonato and Cavalca (2014), seeking the convergence of the contact force and the displacement approach for EHD lubrication. Parameters for nonlinear stiffness and linear damping are obtained. In sequence, Bizarre et al. (2018) evaluated the behavior of the parameters of nonlinear contact force, compared to dry contact for several loads and rotation speeds, pointing to the consistency of pressure distribution and fluid film thickness parameters. In Carrer et al. (2020) a complete rotor-bearing system was modeled, based on the work of (1990) for experimental comparison purposes.

2.2 Rolling element bearings fault

There have been great efforts to model rolling elements bearing, from the development and investigation of mathematical representation to computational tools to understand frequency vibration from the natural motion of REB. The last led to the investigation of the vibration signature, which is also associated with the most common faults in these components. Darlow et al. (1974) addressed the problem of transmission bearing wear causing defects undetected in normal maintenance procedures.

also applied the narrow band envelope analysis, primarily denoted by High Frequency Resonance Technique (HFRT) to the investigation of resonance frequencies of rolling bearings. The fundamental idea behind it is the interaction between fault and surface, generating a short-duration burst, and its energy is dispersed in a wideband frequency. The method consists of band-pass filtering around the excited resonance frequency, followed by amplitude demodulation to identify the fault characteristic frequency and associated harmonics.

The classical work of (MCFADDEN; SMITH, 1984) modeled bearing fault signals and reviews the state-of-art of HFRT to obtain the envelope signal spectrum, stating that while the latter is a well-established procedure, its spectrum features, as sidebands, were yet not fully understood at that moment.

By that time, most identification analyses used the transducer resonance for demodulated frequency as an effective way to improve the signal-to-noise ratio, although it has been reported the downside in the presence of other frequencies, such as pump cavitation, which may mask the REB fault signal (RANDALL; ANTONI, 2011). It is also pointed out that, once this technique assumes sharp impacts in the presence of fault, it has limitations in advanced distributed damage when the fault signal can be submerged in this spectrum background. Moreover, this is still a widely used technique(GUPTA; PRADHAN, 2017).

Tandon and Choudhury (1997) presented an analytical model to estimate the frequency and magnitude within the low-frequency range, resulting from a bearing localized defect. They assume the vibration signal received by the transducer incorporates the race vibrations passed by a transmission media, as the bearing housing. Several impulse shapes were used. Howard (1994) addresses the principles of kinematics, dynamic behavior, and vibration monitoring of REB over the prior years. His work reviews the fundamentals of REB and several applications, as well as procedures for measuring vibration and signal processing.

Those were numerous and significant contributions to the development of techniques for condition identification of REB, from optimal bandwidth selection, to filtering out noise, separating undesirable signal components, and other demodulation techniques. McFadden and

Toozhy (2000) extended the synchronous averaging technique to bearings vibration monitoring, applied only for inner race (IR) defect, resulting in an envelope spectrum that contains only IR characteristic harmonic frequencies and combinations of sidebands modulation at shaft multiples rotation frequency.

Antoni and Randall (2002) proposed a procedure for advanced spectral analysis. Pointed out the differences between gear and REB signals, the first one being simply periodic while bearing signal can induce pseudo-cyclo-stationary signals, which take into consideration the stochastic natures of REB signals, distinguishing it from cyclo-stationary processes, and analyzing the interactions with gear signal. Antoni and Randall (2003) proposed a broad stochastic model for describing and simulating localized fault vibration in REB. It was established the most relevant indicators are the Fourier transform and the power spectral density (PSD) of the squared signal, emphasizing the squaring advantages. In recent work, a pseudo-cyclo-stationary signal model was analyzed, comparing clearance (jitter) and joint slip in the rolling element locations, contributing to diagnostic based on spectral analysis. (BORGHESANI et al., 2022)

One of the proposed methods for enhancing the bearing signal for spectral analysis is the Spectral Kurtosis (SK) (ANTONI, 2006; ANTONI; RANDALL, 2006). The methodology was first introduced by Dwyer (1983) as a statistical tool that indicates the non-Gaussian signal components and the respective location in the frequency domain. SK supports incipient fault identification, based on the aforementioned impulse-like trait in signals with early symptoms of fault. Kurtosis is a statistical moment that represents peak characteristics in a distribution, consequently, SK indicates the ideal frequency band to extract the mechanical signature of the fault. The kurtogram is a representation of the SK in frequency and spectral resolution coordinates and it can provide the optimal parameters to design a band-pass filter (ANTONI; RANDALL, 2006)

Antoni(2007) presented the Fast computation of Kurtogram (FK), to overcome the computation complexity of the spectral kurtogram and make it suitable for online industrial applications. The proposed algorithm has the resolution as a function of levels, stating the frequency band and carrier optimal frequency for the envelope analyses.

Variations of FK or other kurtogram-based tools were presented, contributing to the need for a more generalist approach able to de-noise the REB signal and select the best bandwidth to demodulate and analyze its spectrum. (LEI et al., 2011) claims the shortcoming of kurtogram based on the Short Time Fourier Transform (STFT) or Finite Impulse Response (FIR) filters, can be improved by Wavelet Packet Transform (WPT) based on the Daubechies wavelet. (SINGH; DARPE; SINGH, 2018) proposed to update the Kurtogram to improve sensitivity to shifting the variance of the input signal. The method is known as the Over-Complete Rational Dilation Wavelet Transform (ORDWT) for filter design. The criteria proposed for band selection is a Temporal Energy Operated Auto-correlated Kurtosis (TEO-AK). (LIN; QU, 2000) compare Morlet wavelet decomposition with other decomposition methods, including de-noising rolling bearings and gears fault signals. Qiu *et al.* (2006) investigate its application in REB prognostics. Both works concluded that the Morlet wavelet can be more suited for impulse-like REB initial faults.

Sawalhi and Randall (2008b) compared numerical models and test rig vibration for gear-bearing systems. The REB model included localized faults on the outer and inner race and ball elements. Although the contact model was Hertzian, clearance was considered in the model between the RE and raceways, and also small damping coefficients to represent the dissipative effect of the lubricant film. The identification applied spectral kurtoses for optimized demodulation, and the squared envelope spectrum for those fault locations are presented for simulated and tested data. As an extension, Sawalhi and Randall (2008a) completed the numerical model incorporating the extended fault model, representing rough surfaces, into both the inner and outer races. Due to the small energy dissipated in these conditions, noticeable in the power spectral density, the SK was no longer an option for band demodulation. In both cases, the presence of strong gear-modulated excitations demands filtration and optimum band demodulations to isolate the faults indicative frequencies.

Mishra, Samantaray, and Chakraborty (2017) compared three distinct models of REB and validated them through experimental data. These models include a five Degree of Freedom (DOF) model dry contact; a bond graph model; and a spatial Multi-Body dynamics Simulation (MBS) model in ADAMS® software. The envelope spectrum resulting from different models was compared against experimental data, showing similar frequency bands and sidebands for

MBS and experiments. For faults in the outer race and inner race, all models showed satisfactory results.

Randall and Antoni (2011) present a tutorial on bearings fault modeling and best practices for signals analysis, covering possible steps to a successful envelope spectrum, discrete/random separation, removing effects of small speed fluctuation, and filtration of the optimum band frequency for demodulation. Gupta and Pradhan (2017) condensed recent work on REB vibration analysis, comprehending diverse stages and locations for fault and defects, and their most suitable techniques for fault detection.

Other techniques for faults identification in rolling bearings include decompositions in Intrinsic Mode Frequency (IMFs), based on the Huang-Hilber transform(HUANG et al., 1998) (LEI et al., 2013) reviewed the Empirical Mode Decomposition (EMD) application to the fault diagnosis of rotating machinery, including improvements and combinations with other techniques. Although IMF decomposition, EMD, and Variational Mode Decomposition (VMD) were applied for demodulation or filter-bank in numerous articles. Their advantages on REB diagnoses over STFT and wavelets are debatable (RANDALL; ANTONI, 2023).

2.3 Machine learning fault identification

Recently, many authors have shown automatized methods for fault detection based on machine learning algorithms. Such processes are becoming more popular every day, and as important as choosing the best approach and methodology, feature extraction and selection are some of the key characteristics of this process. Most authors use multi-dimensional statistical features.

Guo *et al.* (2009) performed feature selection from the maximum amplitudes of the envelope spectrum, predefining four characteristic frequencies analytically calculated based on dimensions and operation conditions. This approach is not ideal, since small fluctuations in the rotation or sampling frequency may lead to false information and it is impractical for broad application.

A great number of works for REB fault classification, select statistical data from the time, frequency, and time-frequency domains have been published. All these works apply static and shape attributes of both the time and frequency domain, usually the fast Fourier transform (FFT (NGUYEN; KIM, 2015) of the raw signal. Other features from the time-frequency domain include the application of wavelet kurtogram to evaluate the envelope power spectrum harmonics as features. Other applications use the biorthogonal WPD coefficients as time-frequency domain and perform feature selection to identify the most significant ones.(KIMOTHO; SEXTRO, 2014; LESSMEIER *et al.*, 2016)

Yan and Jia (2018) use the spectrum of IMFs obtained via VMD in combination with other common time-signals-based features. Cui et al. (2022) applied sensitivity analysis to select the features for REB wind turbine applications. Wang *et al.* (2023) proposed a framework method for a digital twin to construct a data model using Pearson correlation to update the model with healthy, faulty, and other simulated numerical data for online information, the features are time-domain based obtained by the simulated twin model.

These studies represent noteworthy efforts to improve methods and practices for REB fault vibration diagnostic. The growth in published articles has been significant, and comprehensive work that collects and reviews all this information greatly contributes to the field. Worden *et al.* (2011) sampled data on the emerging works between 2008 and 2010, and detailed the feature extraction algorithms and practices in machine learning. The author disclosed that the bearings were the most common component for condition monitoring, and Multi-Layer Perceptron (MLP) and SVM were the two most applied algorithms for this purpose. In industrial applications, most Artificial Intelligence (AI) applications combine feature extractors with signal-preprocessing techniques, according to Liu *et al.* (2018) in the review of rotating machine fault diagnosis. Lei *et al.* (2020) overviewed machine fault diagnosis based on machine learning including potential guidelines for future developments.

More specific reviews in REB fault severity assessment were presented Cerrada *et al.* (2018), focusing on relevant techniques and data-driven methods, between 2010 and 2016, to estimate fault size and degradation progress. Hakim *et al.* (2023) through a systematic review of rolling bearing fault diagnoses, covered the main public datasets for bearings fault diagnoses and the most used machine learning and deep learning algorithms.

This literature presented fundamentals for developing this study, and, in light of the latest research, the present ball bearings model is in good agreement with the representation applied in industrial and academic applications. The demand for ML models to supervise health conditions on in-site applications motivates the presented technique as useful for data generation, in a context of deficiency of broad data, in terms of faulty location, size, and general operation conditions for rolling element bearings.

The approach for feature extraction, in many ML classifiers, is essentially rooted in statistical methods based on shape or pattern recognition. In addition, those procedures rely on data that have not been adimensionalized, turning the model specific, and valid only for a particular set of conditions. The proposed method fills the gap in creating broad datasets and extracting information by combining simulated and experimental data.

3 METHODOLOGY

This chapter comprehends the theoretical formulation involved in the bearing motion, the contact theory basis for characterizing the force parameters, how the REB effects can be transported to a system, and by what means this system response can be interpreted and used to create features. The simulated data supply the information to construct broader models for fault classification in real data. The application of signal analyses to treat both simulated and real data is also covered in this chapter.

Vibration is usually measured to be a failure criterion, the normal operation of a rotating system has its response amplitudes and frequencies altered when subjected to any kind of unusual condition. REB already has a cyclic vibration pattern as a result of RE internal arrangement, and when affected by a localized fault, the elements passing through the surface irregularity cause the REB to respond by modulating the input vibration. However, this discrepancy is unclear in some failure stages or can be hidden in the spectrum due to other vibration sources.

Accordingly, the information from time-response, vibration goes through an envelope transformation, Hilbert Transform, and its squared envelope spectrum is encoded to be applied in a Machine Learning classifier algorithm, to identify patterns on the signal that may not be trivial for manual interpretation.

The ML algorithm selected for diagnoses is the Support Vector Machine (SVM) classifier, based on solving an optimization problem, with necessary and sufficient conditions for a global optimal solution, with the model behavior accessible to interpretation.

The analysis methodology shows the careful process of creating representative signals, using the EHD model for the REB contact lubrication, and adding defects in the dynamic model to obtain the time domain vibration response. The method for encoding these signals and choosing the best parameters for the estimator is simultaneously defined by optimizing the training set of samples, and the ML model is tested against unseen data for evaluation. This is performed in a nested cross-validation.

3.1 Ball Bearings Contact Model

The equations presented in this section are derived from the movements observed in a ball bearing, Figure 1, representing the most common type of rolling bearing. Initially, it is assumed dry contact for the load and displacement distribution. Later on, these equations will be extended to incorporate the model with full fluid film lubrication.

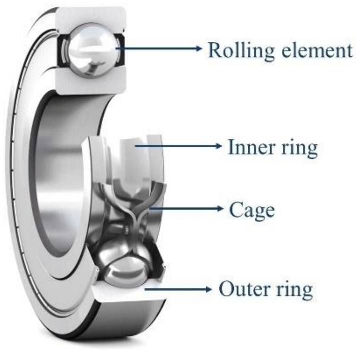


Figure 1. Ball bearing representation, adapted from SKF,(2024)

The first assumption is a bearing under static load, applied on the center of the inner ring by the shaft. This load is divided among the rolling elements placed in the load zone, and those elements are in contact with the inner and outer ring, the deformation and therefore the reaction force are expressed in both contacts.

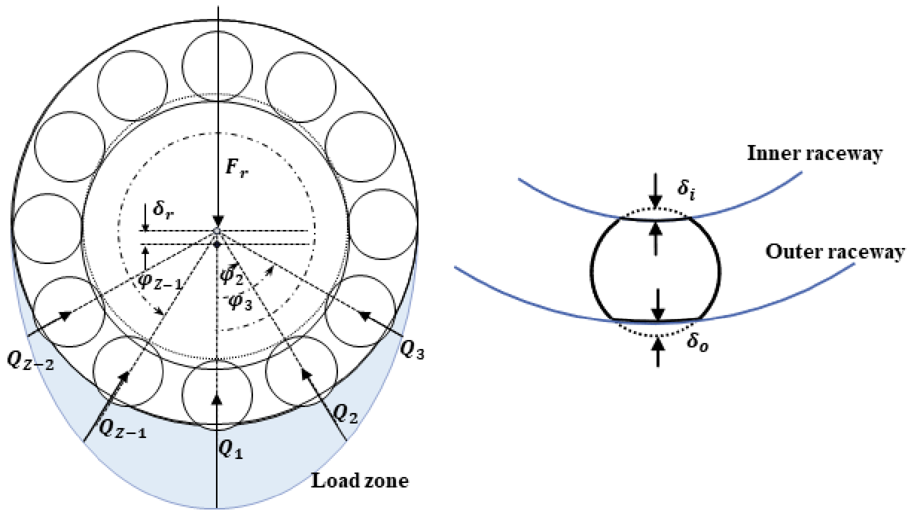


Figure 2. Rolling bearing load distribution and element deformation.

The dry contact can be express through a relation between static load, $Q_{i,o}$, and displacement, $\delta_{i,o}$, by Hertz equation,

$$Q_{i,o} = K_{i,o}\delta_{i,o}^{d} \tag{1}$$

where the inner and outer race, are represented by the subscripts i and o, respectively.

In this model the exponent d is 3/2 for ball bearings and 10/9 for roller bearings, the $K_{i,o}$ is the dry contact stiffness, given by the relationship of the races and ball material, as well as equivalent curvatures radius (SAWALHI, N.; RANDALL, 2008b). The resultant stiffness coefficient, K_H, is valid for all elements positions:

$$K_H = \left(K_i^{-1/d} + K_o^{-1/d}\right)^{-d} \tag{2}$$

The ball is subject to the maximum load, when positioned in line with the applied force F_R, and in this case, when the azimuth angle is zero, it is represented by the Stribeck approximation for a zero clearance and pure radial load.

$$Q_{maxR} = \frac{4.37F_R}{Z\,cos\,\alpha} \tag{3}$$

where Z is the total number of elements in the bearing and α, is the nominal contact angle.

The maximum radial displacement, δ_{max}, can be found through Equations (1) and (3). The displacements of all other balls are determined as a function of the azimuth position and the radial clearance, C_r.

$$\delta_j = (\delta_{max} - C_r)\cos\varphi_j - C_r \tag{4}$$

The first approximation is the displacement and load distribution due to a static load. To represent the dynamic behavior, the gyroscopic moment, M_{gj}, and the inertia are account for each element, and it result in a separate contact angle for the inner race, α_{ij}, and outer race, α_{oj}, as represented in Figure 3, where Q_{oj} and Q_{ij} are the loads supported by each element contacts. The subscript j goes from 1 to Z, representing the number of rolling elements.

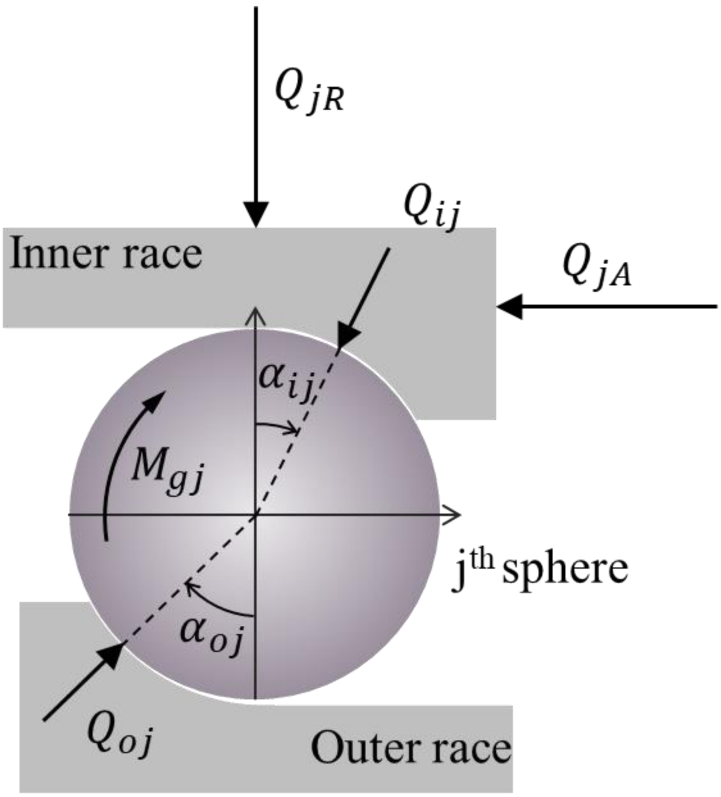


Figure 3. Ball-races dynamic equilibrium.

The dynamic equilibrium on each element is given by:

$$Q_{ij} \sin \alpha_{ij} - Q_{oj} \sin \alpha_{oj} - \frac{M_{gj}}{d_e}\left(\lambda_{ij} \cos \alpha_{ij} - \lambda_{oj} \cos \alpha_{oj}\right) = 0 \quad (5)$$

$$Q_{ij} \cos \alpha_{ij} - Q_{oj} \cos \alpha_{oj} + \frac{M_{gj}}{d_e}\left(\lambda_{ij} \sin \alpha_{ij} - \lambda_{oj} \sin \alpha_{oj}\right) + \frac{\rho \pi d_e^3}{6} d_p \omega_p' \cos \alpha_{ij} = 0 \quad (6)$$

where d_e is the spherical element diameter, and d_p is the pitch diameter, λ_{oj} and λ_{ij} are slip coefficients for both raceways, ρ is the ball specific mass and ω_p is the rotational velocity at the ball center. The complete development can be found in (Radaelli 2013)when applying Equation (1) and the geometrical relation on the bearing equilibrium, there will be dependence of two variables, δ_{ij} and δ_{oj}.

To obtain the adjusted displacement and load in all the contacts between balls and raceways, an encapsulated Newton-Raphson algorithm runs the equilibrium for Equations (7) and (8) to find the pair δ_{ij} and δ_{oj} in each element. Then the bearing equilibrium adjusts the distribution of load on the bearing, as in:

$$F_R - \sum_{j=1}^{Z} \left(Q_{ij} \cos \alpha_{ij} - \frac{M_{gj}}{d_e} \lambda_{ij} \cos \alpha_{ij} \right) \cos \varphi_j = 0 \tag{7}$$

$$F_A - \sum_{j=1}^{Z} \left(Q_{ij} \sin \alpha_{ij} - \frac{M_{gj}}{d_e} \lambda_{ij} \sin \alpha_{ij} \right) = 0 \tag{8}$$

being F_A the trust load on the bearing center.

These relations result in the load distribution and the displacement on each contact being an initial guess in the approximation for the EHD lubrication.

3.2 Elastohydrodynamic Lubrication Contact Model

The lubricant primary function is to reduce friction by avoiding direct metal-to-metal contact, decreasing the risk of wear and corrosion, and therefore prolonging the bearing life. The bearing lubricated contact model must represent not only the heat dissipative component, and viscous damping, but also the oil film thickness, its influence on the bearing load distribution and its vibration response.

The ball bearing with dry contact model is usually represented by Equation (1), the lubricant effect can be roughly approximated by adding preload to simulate the oil film and simply estimating the dissipative force with proportional damping . Although more realistic than just assuming Hertzian contact, these models neglect the complexity of the oil film dynamic behavior.

The interaction among oil film and bearing surfaces cannot be represented only by hydrodynamic lubrication, since this theory omits the deformation of the bodies in contact with the lubrication. High levels of pressure in small or nonconforming geometry areas in the contact lead to Elastohydrodynamic (EHD) lubrication. In a ball bearing, the EHD theory grasps the dynamic behavior of the full film lubricant with the rolling elements and raceway deformation. This theory incorporates the Reynolds equation with the lubricant viscosity-pressure and density-pressure relations, and the elastic deformation in the contact area under pressure. The system of equations integrates both the equation of motion and the static equilibrium of each rolling element.

In Nonato and Cavalca (2014), the solution for the EHD lubricated contact is used to obtain the response to a static load and transient of an initial perturbation. 'The initial estimate in the EHD integration assumes load and displacement distribution as dry contact, as presented in (Radaelli, 2010), adimensionalized as Moes parameters, as described in Venner and Lubrecht (2000). The numerical method to solve the EHD contact is the multilevel finite difference integration, solving the following system of equations, (Nonato, 2011):

- Reynolds equation is a particular case of the Navier-Stokes Equation constraining the differential pressure as positive in the oil film under EHD regime:

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{12\mu}\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{\rho h^3}{12\mu}\frac{\partial p}{\partial y}\right)-u_m\frac{\partial \rho h}{\partial x}-\frac{\partial \rho h}{\partial t}=0 \tag{9}$$

  where p is the contact pressure, h is fluid film thickness, μ is the lubricant viscosity, ρ is the lubricant density and $u_m$is the surfaces relative velocity.
- The elastic deformation in the contact as a function of the pressure distribution.
- Viscosity-pressure relation.
- Density-pressure relation.
- Forces equilibrium across the contact area, considering the pressure distribution.

Considering the static equilibrium of forces, the system integration gives the displacement due to static load in each contact. With the transient response, displacements, velocity, and dynamic load are obtained. This process of integration is computationally expensive. To be suitable for application in complex systems, a reduced contact model of EHD response represents the nonlinear restoring contact force for the bearing characterization.

The contact between the paired surfaces (ball-to-inner race and ball-to-outer race) is assumed to be continuously lubricated with oil, resulting in an elastohydrodynamic (EHD) contact model. The reduced model for EHD contact force is given by the equation proposed in Nonato and Cavalca (2014).

$$\left(Q_{static\,j}\right)_{i,o}=K\left(\delta_j\right)_{i,o}^{d}+\Delta F \tag{10}$$

where K, d and ΔF are the equivalent parameters for stiffness, displacement exponent and offset force, respectively, representing the load, Q_{static_j}, in each element and each contact with the inner and outer raceways.

The Levenberg Marquardt optimization is applied to find these parameters for the distribution of restitutive forces and static displacements from the EHD simulations. The dissipative force and velocities are derived from the transient perturbation method and the damping parameter, D, results from the linear least mean square optimization, where the $Q_{dynamic_j}$ stands for the dynamic load and $\dot{\delta}_j$ is the velocity of the j^{th} element.

$$\left(Q_j\right)_{i,o} = K\left(\delta_j\right)_{i,o}^{d} + \Delta F + D\left(\dot{\delta}_j\right)_{i,o} \quad (11)$$

To obtain the total equivalent parameters of the bearing, it is necessary to decompose the restitutive and dissipative forces of each contact in the axial and radial directions, represented by the subscripts A and R, respectively:

$$\left(Q_{Aj}\right)_{i,o} = \left(Q_j\right)_{i,o} \sin\left(\alpha_j\right)_{i,o} \quad (12)$$

$$\left(Q_{Rj}\right)_{i,o} = \left(Q_j\right)_{i,o} \cos\left(\alpha_j\right)_{i,o} \quad (13)$$

With decomposed loads into radial and axial directions, the total displacement can be found as the sum of internal and external displacements:

$$\left(\delta_j\right)_{A,R} = \left(\delta_{ij}\right)_{A,R} + \left(\delta_{oj}\right)_{A,R} \quad (14)$$

Up to this point, the model comprehends parameters for K, d and ΔF in radial and axial directions, after carrying on the optimization separately for both directions. As the damping parameter, D, is considered linear and the same for both directions:

$$Q_{Aj} = K_A \delta_{Aj}^{d_A} + \Delta F_A + D\dot{\delta}_{Aj} \quad (15)$$

$$Q_{Rj} = K_R \delta_{Rj}^{d_R} + \Delta F_R + D\dot{\delta}_{Rj} \quad (16)$$

This model characterizes parameters for angular contact REB. Nevertheless, the angular contact, alfa, in the following equations, is zero, as the fault bearings data available is from radial REB. The dynamic response of the ball bearing is found through the numerical integration of equations of motion system. The center of the inner ring displacement, relative to the outer ring, which represents all the movement on the bearing node, is denoted by δ_r, and its components in the x-y plane are, δ_{rx} and δ_{ry}. This displacement can be read in the radial direction, for each ball it can be written based on the angular position, the azimuth angle, φ_j, of the *j-th* ball.

$$\delta_j = \delta_{rx} \sin \varphi_j + \delta_{ry} \cos \varphi_j + C_r + C_d \beta \tag{17}$$

the C_r represents the radial clearance, C_d is the fault or defect maximum depth, β is the fault operator that defines whether or not the element is passing through the fault. The azimuth position, φ_j, depends on the nominal cage rotational speed, ω_c, equivalent to the center of the sphere orbit

$$\omega_c = \frac{\Omega}{2}\left(1 - \frac{d_e}{d_p}\cos \alpha\right) \tag{18}$$

$$\varphi_j = \mathrm{remainder}\left(\frac{\omega_c t}{Z}(j-1), 2\pi\right) \tag{19}$$

Ω is the inner ring rotational speed, considering the outer ring fixed, d_e and d_p, are the ball and pitch diameter, respectively, and α is the contact angle. the function remainder (u, v) gives the remainder after the division of u by v.

Deriving the ball radial displacement, Equation (17), to obtain the radial velocity of the j^{th}ball, $\dot{\delta}_j$, where the $\dot{\delta}_{rx}$ and $\dot{\delta}_{ry}$ are the components of the velocity of the center of bearing inner ring.

$$\dot{\delta}_j = \dot{\delta}_{rx} \sin \varphi_j + \dot{\delta}_{ry} \cos \varphi_j + \omega_c\left(\delta_{rx} \cos \varphi_j - \delta_{ry} \sin \varphi_j\right) + C_d \dot{\beta} \tag{20}$$

the $\dot{\beta}$ is the operator β time derivative, representing the velocity of the ball going into and out of the spall and it represents the fault excitation on this element.

To complete the displacement and velocity of each element, it is necessary to cover the influence of the spall on each REB component. Three different types of faults are considered in this model, inner race, outer race, and ball element fault. In these cases, the fault is characterized as a discontinuity of a depth C_d at the initial angular position θ_{d0}, along an angular span of $\Delta\theta_d$.The subscript d denotes the fault-related parameters.

Outer race fault

In this case, each sphere, during a complete cycle around the bearing center, will lose contact with the surface once into the spall. Therefore, in a complete shaft rotation, the spall is hit by the exact number of spheres, Z, in the bearing. As a result, the characteristic frequency of this type of defect is the Ball Passing Frequency in the Outer ring (BPFO):

$$BPFO = \frac{Z\omega_c}{2\pi} = Z\frac{\Omega}{4\pi}\left(1 - \frac{d_e}{d_p}\cos\alpha\right) \tag{21}$$

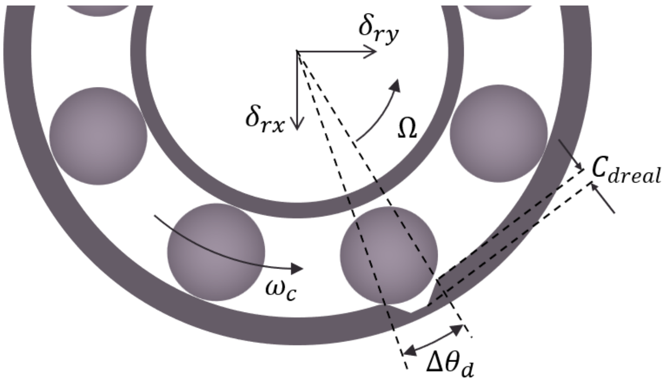


Figure 4. Outer race fault.

Figure 4 represents the RE passing by a spall in the outer race, with $\Delta\theta_d$ as the angular fault span. The expression for the displacement due to the fault, δ_{dj}, depends on the β operator, which defines whether the element is passing by the fault or not. When the spall is incipient, β is closed to a step function, since the RE rapidly passes through the discontinuity. As the fault increases, a smoother function is proposed to represent the spall path.

$$\beta = \begin{cases} f(\varphi_j), & \theta_{d0} < \varphi_j < \theta_{df} \\ 0, & otherwise \end{cases} \tag{22}$$

In Equation (22), θ_{d0} and θ_{df} are, respectively, the initial, subscript 0, and final, subscript f, angular position of the spall, with:

$$\theta_{df} = \theta_{d0} + \Delta\theta_d \tag{23}$$

The function $f(\varphi_j)$ defines the smooth profile of the RE pathway throughout the spall region. This function must represent the RE exiting the of spall mirroring the spall entering, In addition, $f(\varphi_j)$ must start and end in zero, the normal path, as well as having its maximum depth in the half of the way. The angular displacement of the RE relative to spall entry is $\varphi_j - \theta_{d0}$, hence the profile can be written as a function of the RE angular position, φ_j, shifted by θ_{d0} and expanded in terms of $\frac{\pi}{\Delta\theta_d}$.

$$f(\varphi_j) = 1 - \left|\cos\frac{\pi(\varphi_j - \theta_{d0})}{\Delta\theta_d}\right| \tag{24}$$

In the same way, the velocity operator is given by $\dot{\beta}$, that imposes velocity to the RE in the radial direction, and is given by:

$$\dot{\beta} = \frac{d\beta}{dt} = \begin{cases} \dot{f}(\varphi_j), \; \theta_{d0} < \varphi_j < \theta_{d0} + \Delta\theta_d \\ 0, \; otherwise \end{cases} \tag{25}$$

$$\dot{f}(\varphi_j) = \mathrm{sign}\left(\cos\frac{\pi(\varphi_j - \theta_{d0})}{\Delta\theta_d}\right)\frac{\pi}{\Delta\theta_d}\omega_c \sin\frac{\pi(\varphi_j - \theta_{d0})}{\Delta\theta_d} \tag{26}$$

The $f(\varphi_j)$ contains an absolute value function and, although the ordinary derivative does not exist at the turning point, its symmetric derivative is zero (THOMSON, 1994). This assumption represents the behavior of the RE when reaching the lowest point and momentarily stopping when forced to change direction and exit the fault. This is expressed by the function sign(·).

This behavior is modeled as an excitation of forced displacement and velocity on the element. This is exemplified in Figure 5 (a), where the dashed green line is the step function usually applied, and the cyan line is the smooth β function. Figure 5 (b) represents the velocity excitation of the fault corresponding to $\dot{\beta}$, and the discontinuity is a result of the change in the path orientation when the ball achieves the bottom of the fault.

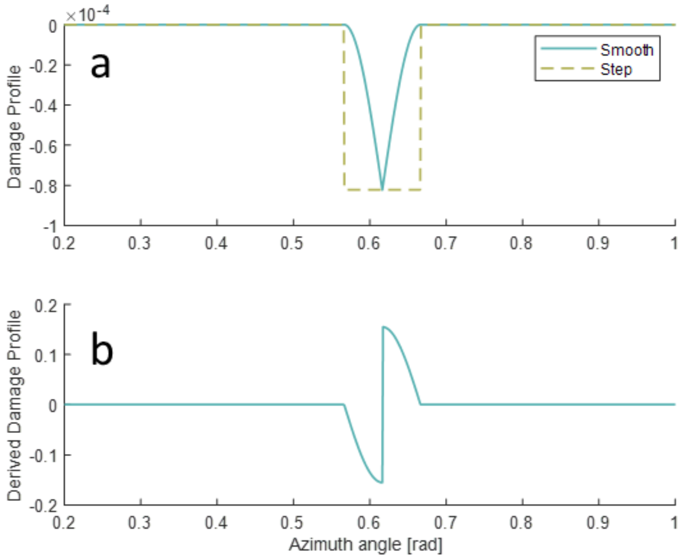


Figure 5. Damage profile (a) displacement and (b) velocity.

If the ball loses contact with the inner race, as a consequence of the spall depth, it will no longer participate in the dynamic load distribution until leaving the spall. In that case, the discontinuity of $\dot{\beta}$ and the cusp in the β would not be accounted. Otherwise, if the fault is not deep enough compared to its length and the ball hits the bottom, the discontinuity would act as a result of this impact. If the fault depth happens to be greater than the inner race radial displacement, the element will lose contact and no longer participating in the load distribution, as exemplified in Figure 6.

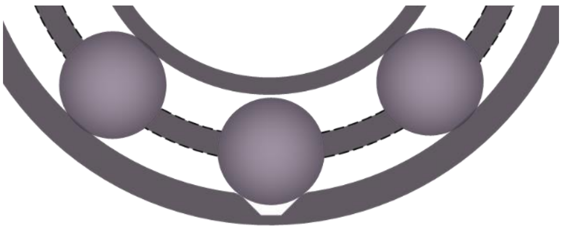


Figure 6. Unloaded rolling element when passing through the spall.

In this case, Figure 6, there is a transient when the RE restores contact with the inner race, as if a spring damper is added in parallel with the other REs, altering the load distribution once again.

<u>Inner race fault</u>

For the inner race, the fault behavior is similar to the outer race, except that the spall rotates along the inner race and the shaft, leading to a different response and characteristic frequency.

The spall rotates in the same orientation as the elements do. The relative angular velocity between spall and elements is the difference between the shaft rotational velocity, Ω, and the cage velocity, ω_c.

Considering the number of times the fault collides with an RE, the frequency is multiplied by the number of RE in the bearing, Z, resulting on the Ball Passing Frequency in the Inner ring (BPFI), as represented in Figure 7.

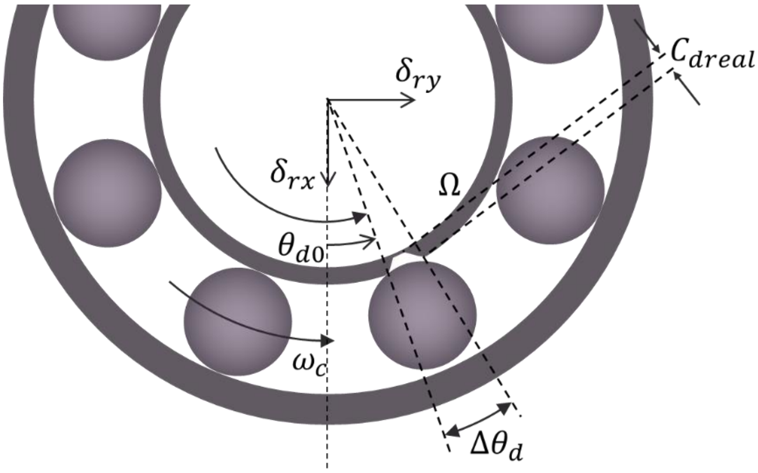


Figure 7. Inner race fault.

Therefore, the BPFI, is the characteristic frequency related to this defect.

$$BPFI = \frac{Z(\Omega - \omega_c)}{2\pi} = Z\frac{\Omega}{4\pi}\left(1 + \frac{d_e}{d_p}\cos\alpha\right) \tag{27}$$

The representation of the spall in the inner race is very similar to the outer race, except for the spall angular position. As the inner race rotates along with the shaft, with rotating speed Ω, the spall angular position rotates along with it.

$$\theta_{d0} = \text{remainder}\ (\Omega t, 2\pi) \tag{28}$$

$$\theta_{df} = \text{remainder}\ (\Omega t + \Delta\theta_d, 2\pi) \tag{29}$$

Equations(22), ((*22*) and (25) are applicable in the same way as in the outer race fault, and the function of the profile derivative in the fault region is given by:

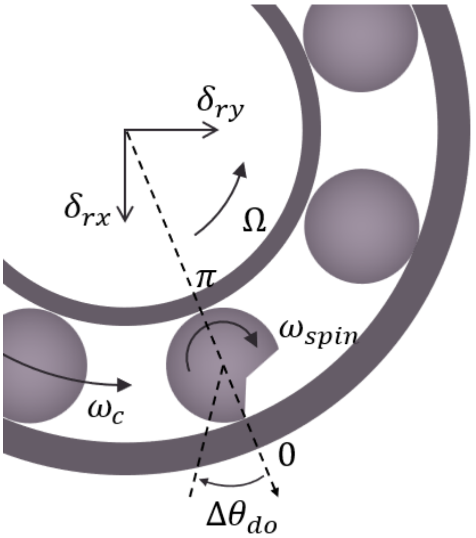
$$\dot{f}(\varphi_j) = \mathrm{sign}\left(\cos\frac{\pi(\varphi_j - \theta_{d0})}{\Delta\theta_d}\right)\frac{\pi(\omega_c - \Omega)}{\Delta\theta_d}\sin\frac{\pi(\varphi_j - \theta_{d0})}{\Delta\theta_d} \quad (30)$$

An observable effect is the variation of intensity in the spall and RE collision, when the spall goes through the load zone, the bearing reaction force is more responsive to the discontinuity. Thus, this fault results in a modulation by system rotational frequency.

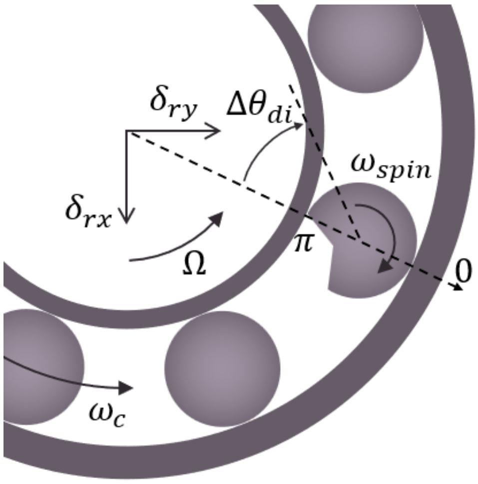
Ball fault

The spherical rolling elements motion defines the radial displacement and velocity bearings, representing the inner and outer race relative movement, δ_r and $\dot{\delta}_r$. The elements that orbiting around the bearing center are assumed to have the same frequency as the cage, ω_c, besides their spin rotations. The last one is responsible for characterizing the fault in the rolling element (RE) surface.

A spall in the RE surface spins around its rotational axis, hitting both races in every complete spin, leading to a higher impact when going through the load zone, resulting in a characteristic frequency two times the spin, and modulated by the cage frequency.



a



b

Figure 8. Ball fault going through (a) the outer race and (b) the inner race.

To obtain the spin rotation, it is considered that there is no slippage in the contact point in both races. To determine the ball-races dynamic contact, the reference frame revolves around the bearing center, as represented in Figure 8.

$$\omega_{spin} = \omega_c \left(\frac{d_p}{d_e} - \cos\alpha\right) = \frac{\Omega}{2}\frac{d_p}{d_e}\left(1 - \left(\frac{d_e}{d_p}\cos\alpha\right)^2\right) \tag{31}$$

The position of the spall in the sphere surface is a time-function and depends on the initial position of the spall, θ_{d0}. The spin angle, θ_s, is considered negative, since it spins in the opposite direction as the cage and inner race.

$$\theta_s = -\omega_{spin} t + \theta_{d0} \tag{32}$$

Since the spall angular span, $\Delta\theta_d$, is determined in terms of the ball dimension, it depends on what race the spall is in contact with. If it is passing through the inner race, $\Delta\theta_{di}$, and if the span is in the outer race, $\Delta\theta_{do}$, in the spin frame of reference, shown in Figure 9:

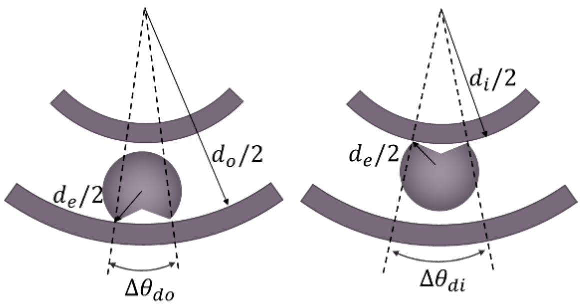


Figure 9. Detail on of the ball fault angular span.

$$\Delta\theta_{di} = \Delta\theta_d \frac{d_e}{d_i} \tag{33}$$

$$\Delta\theta_{do} = \Delta\theta_d \frac{d_e}{d_o} \tag{34}$$

In this way, it is possible to establish the depth of the spall in contact with the inner race, C_{dbi}, and the depth in contact with the outer race, C_{dbo}. Due to differences in the curvature in the inner race, the interaction with the spall is longer and deeper than in the outer race.

$$\beta = \begin{cases} 1, & 0 < \theta_s < \Delta\theta_{do} \text{ and } j = k \\ C_{dbi}/C_{dbo}, & \pi < \theta_s < \pi + \Delta\theta_{di} \text{ and } j = k \\ 0, & otherwise \end{cases} \tag{35}$$

with k being the damaged rolling element.

<u>Maximum possible fault depth</u>

The depth of the spall is constrained by the extension of its angular span and the bearing geometry, and the value measured is only valid as maximum. The fault extension, Δs, is determined based on the location of the fault, where k is the index for the location of the fault, inner race fault (i), outer race (o) or element fault (e), hence the $\Delta\theta_d$ relative to that surface.

$$\Delta s_{\mathrm{k}} = \mathrm{r_k} \sin \Delta\theta_d \cong \mathrm{r_k}\, \Delta\theta_d \tag{36}$$

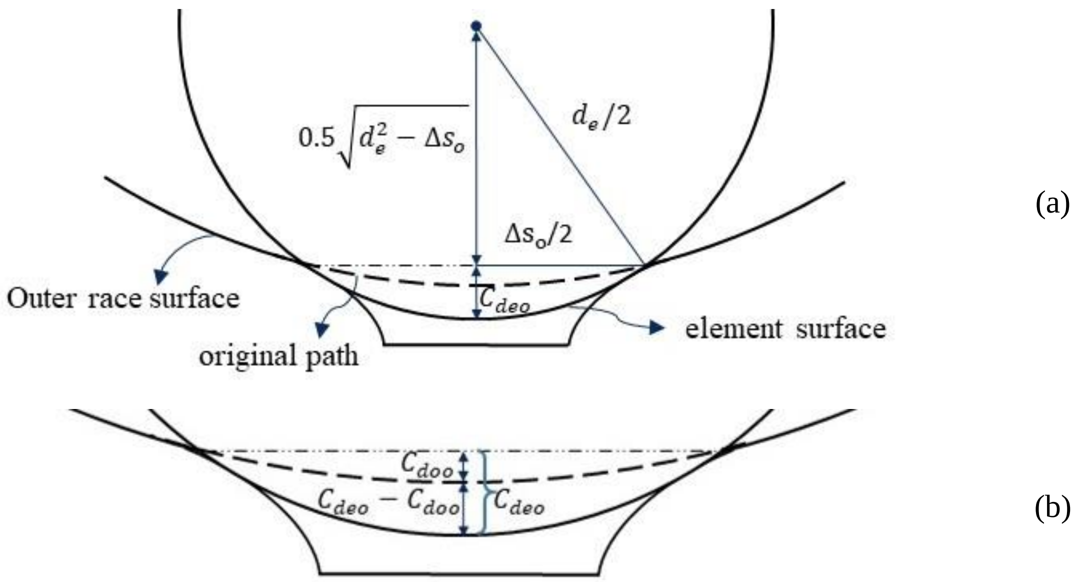


Figure 10. (a) Dimension of effective fault depth (b) surface detail.

Figure 10 represents the curvatures of the contact between the ball and the outer race. The depth relative to each fault location can be achieved as C_{di} is the maximum depth reached in the inner race; C_{de} is the maximum contact loss in the rolling element contact along the spall, C_{do} is the maximum depth in the outer race.

$$C_{di\mathrm{k}} = r_i - \sqrt{{r_i}^2 - \Delta s_{\mathrm{k}}^2} \tag{37}$$

$$C_{de\mathrm{k}} = r_e - \sqrt{{r_e}^2 - \Delta s_{\mathrm{k}}^2} \tag{38}$$

$$C_{do\mathrm{k}} = r_o - \sqrt{r_o^2 - \Delta s_{\mathrm{k}}^2} \tag{39}$$

The maximum depth possible, given the geometry restriction is the minimum between real depth and geometry, differs from the inner race and outer race by the differences in curvature. For outer race fault the depth C_d is the minimum between the difference $C_{deo} - C_{doo}$ and the real measured depth, C_{dreal}

$$C_d = \min(C_{deo} - C_{doo}, C_{dreal}) \tag{40}$$

For inner race fault, due to the curvatures being in opposite position, the total depth is a sum of both curvatures, $C_{dei} + C_{dii}$.

$$C_d = \min(C_{dei} + C_{dii}, C_{dreal}) \tag{41}$$

If the fault is located in the rolling element, the standard value for the maximum depth is the fault in contact with the outer race, and the equation is similar to Equation (40).

$$C_d = C_{dbo} = \min(C_{dee} - C_{doe}, C_{dreal}) \tag{42}$$

And if it is contact with the inner race the maximum depth is defined as:

$$C_{dbi} = \min(C_{dee} + C_{die}, C_{dreal}) \tag{43}$$

Bearing reaction forces

Each RE displacement and velocity, Equations (17) and (20), are restrained to positive values only, as a consequence of two assumptions. First, the REs inertia is neglected and the REs are modeled by a damper-spring system connecting the inner and outer race, by a non-linear reaction force, as in Equation (16). Second, these REs are not coupled to any of the raceway surfaces, so the REs represent a spring damper that only pushes but does not pull. Therefore, it is only relevant to the system the displacement, δ_{Rj}, when the RE is compressed, and the velocity, $\dot{\delta}_{Rj}$, when the inner and outer races are approaching one another.

$$\delta_{Rj} = \max(\delta_j, 0) \tag{44}$$

$$\dot{\delta}_{Rj} = \max(\dot{\delta}_j, 0) \tag{45}$$

The total reaction force is the sum of the reactions of all the elements, from 1 to Z. The total reaction on the center of the bearing is represented as f_{bx} and f_{by}, for the radial directions, by applying Equation (16) on each RE, and then adding their projection in x and y directions, respectively.

$$f_{bx} = \sum_{j=1}^{Z} Q_{Rj} \cos \varphi_j \tag{46}$$

$$f_{by} = \sum_{j=1}^{Z} Q_{Rj} \sin \varphi_j \tag{47}$$

The total reaction forces of the bearings are then added to the rotor system of equations in the respective bearings nodes. The model and degrees of freedom depend on the rotor-bearing-configuration and the location of the acquisition sensors.

3.3 Signal Analysis

Vibration can be considered a failure criterion since it hinders the normal operation of the rotating system. Rolling bearings are subject to surface fatigue failure due to the small contact area that occurs between the rolling element and the inner and outer races, but just the natural movement of the bearing can generate vibrations and noise when subjected to high rotations.

The nature of the bearing cyclical movement influences the rotor signal, in that way, the BPFO modulates the signal, as can be observed in Figure 11, triggered by variations in the elements position during movement when the bearing configuration changes.

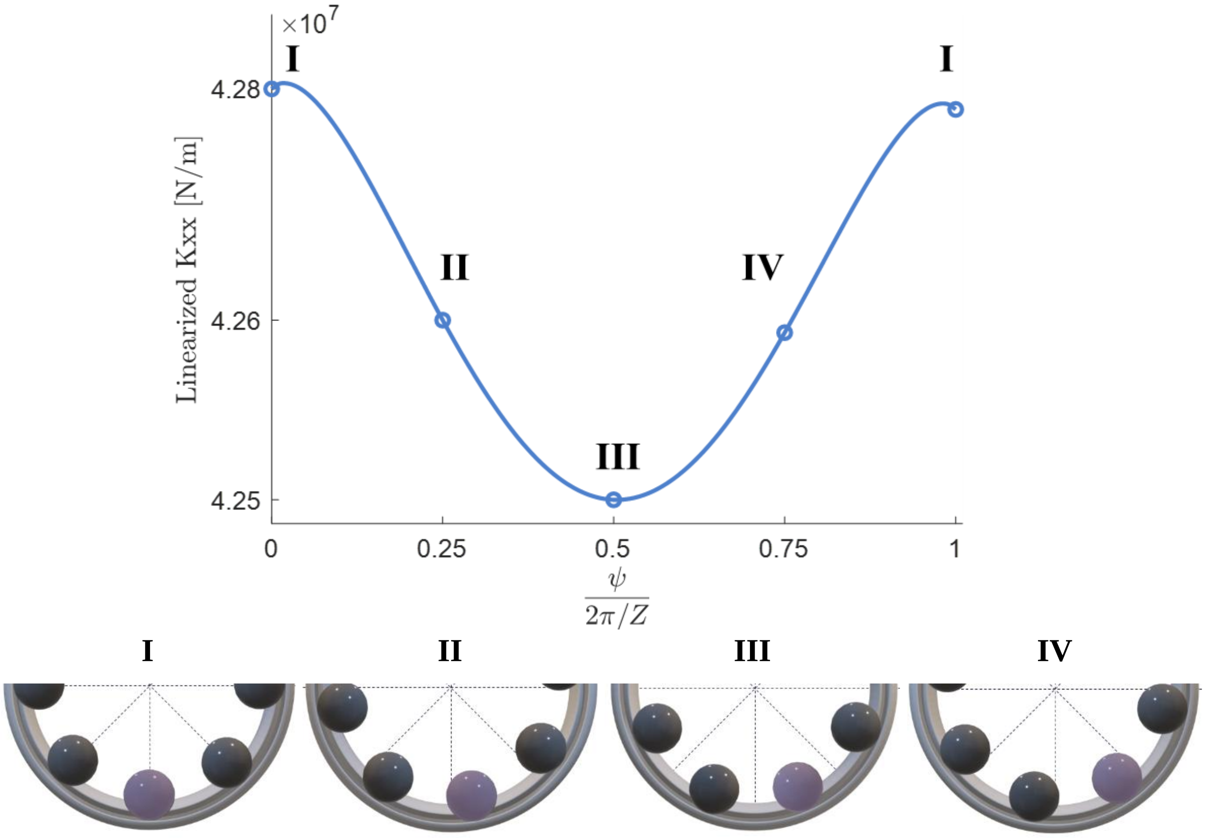


Figure 11. Stiffness variation during a BPFO period.

In addition to this intrinsic condition, there are ideally periodic excitations arising from faults in the internal surfaces. This signal cannot be fully represented by Fourier analysis, directly from the time sequence. The modulation of the rotor signal by the cyclical behavior of bearing, as well as periodic impacts from faults, can be characterized as multiplications rather than a sum of harmonics, with frequencies expressed as multiple combinations. It can be stated that, in the initial stages, those faults generate a train of impulses in time that excites the system, this is represented by the function β, adjusting the fault depth contribution in the total displacement along with the fault angular position.

There are two scenarios, one is the signal modulation by the variant compliance of the bearing, it is indeed a signal multiplication, as exemplified in Figure 11. Another case is the presence of fault that expresses itself in the signal as the sum of periodically spaced impulses, the period of this occurrence is related to the fault frequencies we aim to identify.

Fourier analysis consists of decomposing a signal as a sum of harmonics with amplitudes and phase constant throughout the signal interval. In unconforming bearings, the impulse train produced by the fault leads to instant variations of amplitude and frequency, once the system convolution to the fault input will lead to a transient attenuation back to the prior condition. This cyclic fault impact appears as a modulation of the original signal. One of the main techniques to detect fault frequencies is envelope analysis, especially through Hilbert Transform (HT).

The HT returns the original time series analytic function. Analytical functions are defined as differentiable complex functions whose derivatives are also complex. However, to be valid, the limits of both sides approaching zero must converge to the same solution. Therefore, the function must satisfy the Cauchy Theorem, that is, for the complex number $z = x + \mathrm{i}y$ there is the correspondent analytical function $h(z)$, and, for the function $h'(z)$ exists and be complex, the following relation must be true (WARD BROWN; CHURCHILL, 1996)

$$h(z) = u(x, y) + \mathrm{i}v(x, y) \tag{48}$$

Cauchy Theorem states that:

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} \text{ and } \frac{\partial u(x, y)}{\partial y} = -\frac{\partial v(x, y)}{\partial x} \tag{49}$$

The condition that satisfies Equation (48) is that both real and imaginary parts of $h(z)$ are harmonic functions(OPPENHEIM; SCHAFER, 2010). Another conclusion that can be drawn from this observation, is that the real and imaginary parts of $h(z)$ must be an even and an odd function, respectively, comparable to Euler equations, where the cosine and sine fulfill this premise.

The Fourier transform of a real signal is the real part of a Fourier transform of a complex signal, which is essentially an even function. In this way, to reconstruct the complex signal that satisfies an analytical function condition, it is only necessary to know its real part.

$$x_R[k] = \mathrm{Re}(x[k]) \leftrightarrow X_R[n] = \mathrm{Re}(X[n]) \tag{50}$$

Since any signal can be decomposed in even and odd signals, and in case the signal $x[k]$ be causal, that is $x[k]$ = 0 for $k < 0$, it can be reconstructed only knowing its even component.

$$x[k] = x_{even}[k] + x[k] \tag{51}$$

$$x_{even}[k] = \frac{x[k] + x[-k]}{2} \qquad 52$$

Let $X[n]$ be the FFT of signal $x[k]$ with length L. The analytical signal is obtained by the inverse FFT of the $H[n]$, obtained by multiplying the first half of the $X[n]$ by 2, the second by zero, and the overlap points, by one, that is, n equals to 1 and L/2. (OPPENHEIM; SCHAFER, 2010)

$$H[n] = \begin{cases} 0, & \text{if } L/2 + 2 \leq n \leq L \\ 2X[n], & \text{if } 2 \leq n \leq L/2 \\ X[n], & \text{if } n = 1, L/2 + 1 \end{cases} \qquad (53)$$

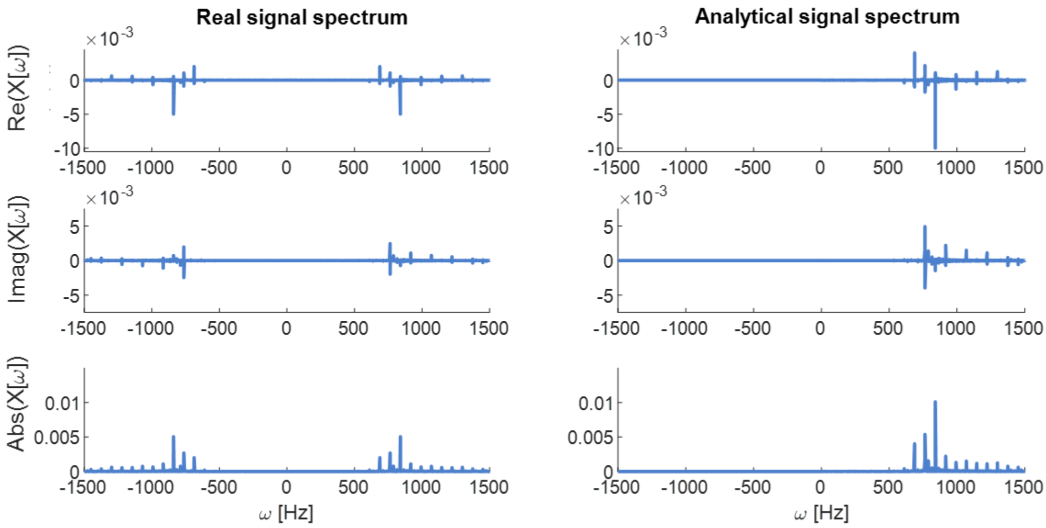


Figure 12. Real and analytical spectrum.

Figure 12 shows the decomposition of a signal into odd and even components, by separating the spectrum into real and imaginary parts. Figure 13 exemplifies a vibration signal and the correspondent envelope containing characteristic frequencies, after performing the HT. The envelope spectrum (ES) is obtained by the Fourier transform of the analytical signal magnitude.

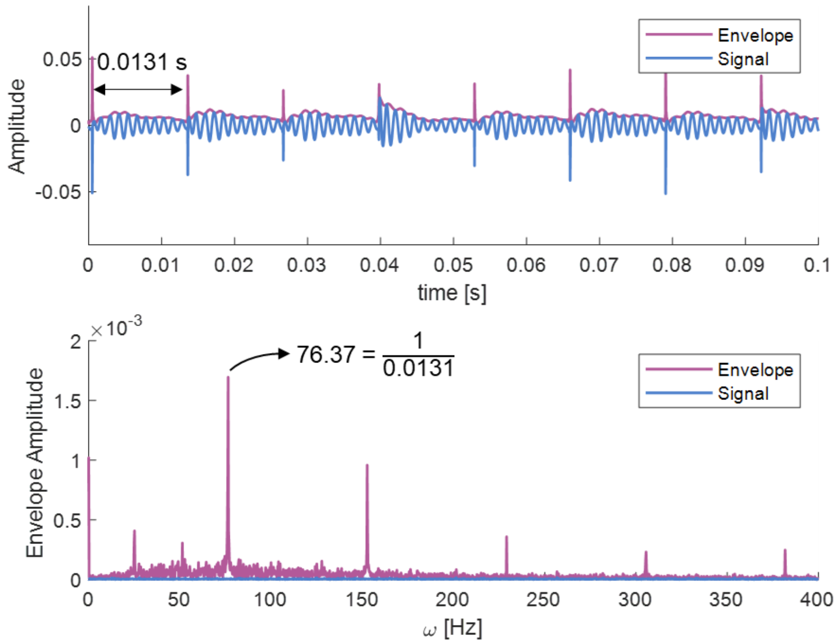


Figure 13 Comparing signal and envelope time amplitude and spectrum.

It is possible to filter the vibration to extract frequency contents where the envelope is more noticeable. One way of doing is by performing a fast kurtogram (FK) (RANDALL; ANTONI, 2011) using filter banks to separate the combination of carrier frequency and bandwidth to estimate the sharpness of the impulses. For each frequency band, the kurtosis is computed and the one with the higher value is associated with more information about failure. One drawback to filtering the optimum band is FK sensitivity to large random impulses or other components with cyclical excitation. Those factors can suppress the fault information, leading to improper frequency range selection.

3.4 Machine Learning

The fault identification technique to be tested in this research relies on Machine Learning (ML). The machine learning process starts selecting the samples for the problem. The source of the samples will dictate the generalization of the final model. The samples must be fully separated in training and test sets. In this stage, any information on the test set must be kept aside to avoid data leakage, that is, the use of any information from the test set for training the model. This means the process of transformation, averaging, filtering should consider the training parameters only.

Three different tests are performed, the first one aims to reproduce the work (LESSMEIER et al., 2016)but with the feature encoder proposed in this work using a similar division of nested cross-validation per groups method, the second one is training and validating the model only with simulated fault signals and one real health bearing dataset. The model is tested with all of Paderborn real fault data sets. The last test is a combination of the previous ones, and consists of evaluating real faults data set augmentation with simulated fault samples.

The samples are divided into training and test sets, and the training samples are divided once again into training and testing sets in an inner loop, for the hyperparameters tuning, performed by the optimization. Henceforth, the inner loop training set division will be referred to as inner training and validation sets.

3.4.1 Nested cross-validation – Real data

This section aims to compare the outcome of this work feature extraction and selection model with the same group-folded nested cross-validations in (LESSMEIER et al., 2016)Each sample is generated from a unique time signal and grouped by each real bearing test rig run. Cross-validation is performed in two steps to define the best parameters and evaluate the model, as represented in Figure 14. In each outer loop fold two groups are separated as the test set and three for the training set. Inside the inner loop, two groups are kept for inner training and one for validation.

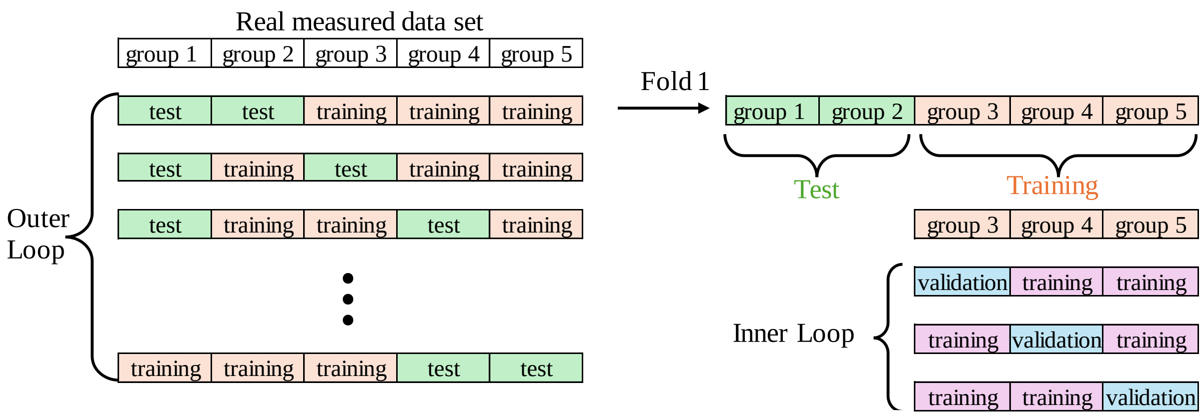


Figure 14. Nested cross-validation example.

Each group contains an equal number of samples for each label, one healthy bearing, and two bearings with real faults, one with inner race and the other with outer race fault. The bearings damages were caused by an accelerated life technique described in (LESSMEIER et al., 2016) The number of samples in each class is balanced, so the algorithm does not favor one class over the others. Samples of the same bearing are not shared between groups, to avoid unrelated to fault pattern recognition. In the inner loop, the same separation in groups is done to prevent overfitting.

Two additional tests are carried out, one training with four groups is testing with just one, to see how each group responds individually to a large number, and variety, of training samples. The other test is how the model behaves when there is just one group of samples available for training the model.

3.4.2 Training set with simulated faults

To successfully model fault identification, enough data is required, especially from diverse groups, which implies different operational conditions or setups. This is difficult to obtain from real data, specifically for various types and sizes of faults.

To build a comprehensive data set, it is possible to manufacture artificially damaged bearings, which would also be expensive and time-consuming, or even unachievable for most in-site machines. A solution proposed in this research is to simulate bearing response in the presence of fault, in as many configurations as intended for training the model with a combination of simulated fault signals and healthy measured signals.

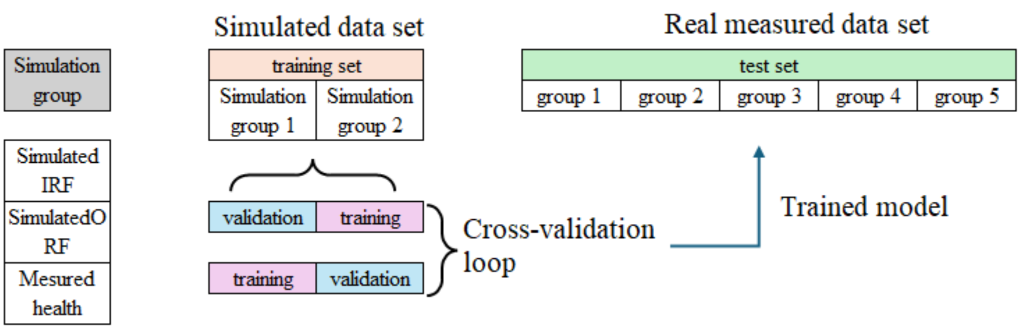


Figure 15. Cross-validation with simulated fault bearings signals.

Figure 15 shows the cross-validation with the simulated fault data set, divided into two groups, with differences in fault size to emulate distinct groups of bearings, the groups also have individual noise levels and unbalanced excitation. Each training group contains one healthy bearing dataset withdrawn from the test set. These healthy bearing signals are additionally applied to add current noise to the simulated signals dataset. All group samples used in the training set are kept completely apart from the test set.

3.4.3 Training data augmentation using simulated faults

This section proposes combining the same simulated data set from section 3.4.2 in addition to one group from the data set of section 3.4.1, in a nested cross-validation where only the real fault data is used for the test. The method is described in Figure 16, the simulated training set is never part of the test set, but in the training set all three groups are used for inner training and validation.

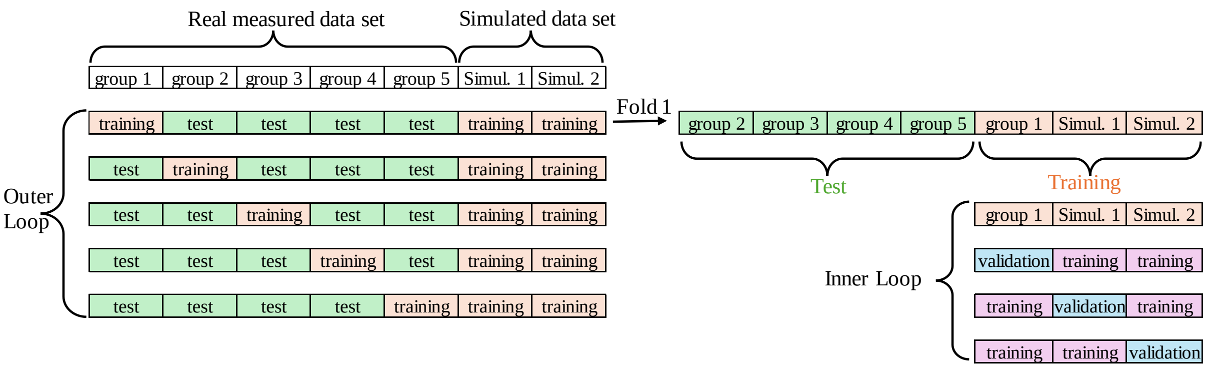


Figure 16. Train-validation-test nested cross-validation with simulated data.

This test aims to evaluate how a simulated data set can improve an ML model that already contains real measured samples but with low variability. This result will be compared with the model trained with just one group of samples.

3.5 Support Vector Machine

There are several algorithms available to build a classification model. However, for this work, a support vector machine is selected based on its known robustness, interpretability of results, and foundation in algebra and optimization theory. These characteristics are desirable to evaluate the feature extraction method.

Support vector machine is a well-known method for supervised machine learning. Developed (VAPNIK, 1995), it consists of solving a non-linear optimization problem of the convex objective function Φ:

$$\begin{cases} \min\limits_{w,b,\xi} \Phi = \frac{1}{2} w^T w \; + C \sum\limits_{i=1}^{N} \xi_i \\ \text{subject to } (w^T \phi(x_i) + b)\mathrm{d}_i \geq 1 - \xi_i \\ \xi_i \geq 0 \\ i = 1,2, \dots, N \end{cases} \tag{54}$$

where ξ_i is a variable for outlier control, C is a hyperparameter that penalizes those outliers, b is the function bias, while $\mathrm{d}_i \in \{-1,1\}^N$ determines the boundary side of the hyperplane, and w is the weight vector. The sample on or within the margins is defined as the i^{th} support vector, x_i, although what is actually considered is the function $\phi(x_i)$ that maps the support vectors in a feature space (FS).

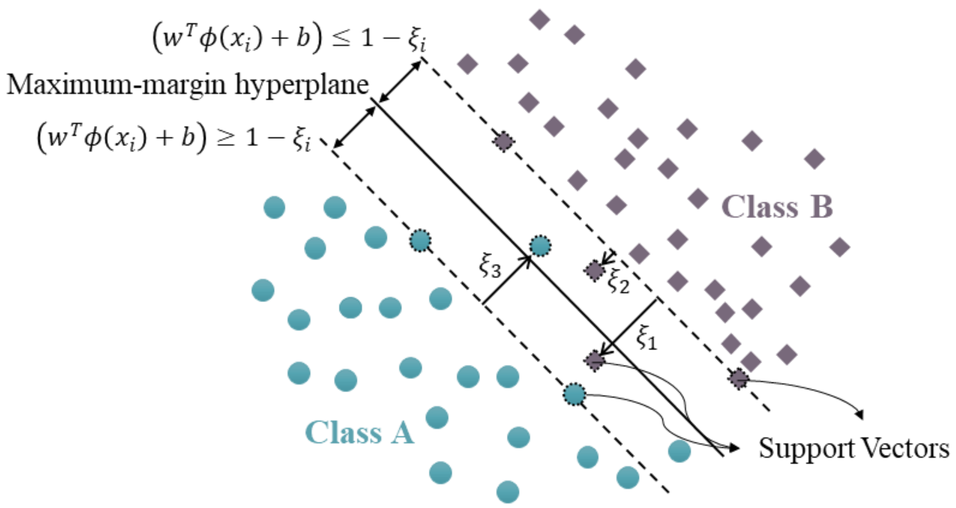


Figure 17. Two-class SVM hyperplane separation.

Figure 17 illustrates how the method separates the data into specific classes by maximizing the margin distance between the adjacent vectors in each class, denominating then support vectors. In the optimization process, they consist of non-trivial solutions and, most of the time, incur a penalty when a sample is misclassified or within the margin boundary. The data fully inside the margin, namely, everything but the support vectors, can fluctuate under the condition of not surpassing the decision boundaries, in a way the result is independent of this remaining data (BISHOP, 2006)

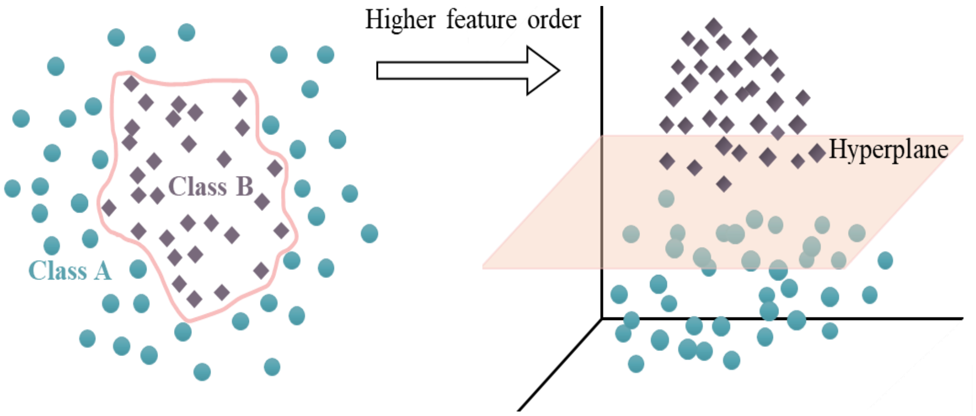


Figure 18. Visualization of two class set separations in feature space.

Figure 18 demonstrates when two different classes cannot be fully separated by a hyperplane and the solution is mapping the constraints, x_i, in a higher dimension FS, $\phi(x_i)$, of unknown order. For the optimization, it is only necessary to know the scalar product of the mapping function $\phi(x_i)$ and the result in the input space is a Kernel function, $\mathrm{K}(x_i, x_j)$, conditioned to satisfy Mercer's Theorem(CORTES; VAPNIK, 1995).

The optimization solves the dual problem to the primal problem in the Equation (54) and the constraints have their inner product replaced by the kernel function, $\mathrm{K}(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ and $\mathrm{K}_{ij} = \mathrm{d}_i \mathrm{d}_j \mathrm{K}(x_i, x_j)$ the σ are the dual coefficients, upper-bounded by the penalty hyperparameter C.

$$\begin{cases} \min\limits_{\sigma} \dfrac{1}{2} \sigma^T \mathrm{K} \sigma \ + \epsilon^T \sigma \\ \text{subject to } \mathrm{d}^T \sigma = 0 \\ \qquad\qquad 0 \leq \sigma \leq C \\ \qquad\qquad i, j = 1,2, \dots, N \end{cases} \tag{55}$$

where ϵ is a vector of ones.

In the case of separable classes by the hyperplane in the input space, the Kernel is the inner product of the support vectors, as the features space and the input space are the same. For all other cases, the Kernel function must be selected to capture the non-linear relationships of the data and to establish clear decision boundaries. Two categories of function are tested for this problem, the radial basis function (RBS) and the polynomial function. The RBS is a Gaussian-shaped function, in which the parameter γ defines the wideness of the function.

$$K_{RBF}(u,v) = e^{-\gamma \, \|u-v\|^2} \tag{56}$$

and the polynomial kernel function, which also depends on the γ, in addition to the degree parameter, d, and r, the linear coefficient.

$$\mathrm{K}_{POLY}(u,v) = (\gamma\langle u,v\rangle + r)^d \tag{57}$$

The type of function as well as its associated hyperparameters, namely γ, d, r and the penalty for outliers C, must be optimized with the problem feature parameters to obtain a model specifically designed for fault classification.

3.6 Feature extraction

Several available publications employ features that are kept dimensionalized, applicable for only specific operation conditions in which the ML model was trained. In some of them, the model was trained and tested by samples of the same time signal and the final model is quite specific, leading to the accuracy biased toward optimism (RAUBER *et al.*, 2021). Fault simulation to construct training features is proposed in this work to solve these shortcomings, which usually result from a lack of training data with variability.

The vibration responses are used as the training set. At this point, it is necessary to determine how broad or specific the model should be, it is a tradeoff, which can compromise the accuracy of the classification. To fully separate the influence of the fault from other components present in the signal can be challenging, as the amplitudes are sensitive to factors such as load, speed, presence of imbalance, misaligns, and even temperature. As a consequence, adimensional features bring less information about fault if we are under a restricted scenario but generalize better for a variety of operational conditions.

Although statistical parameters are extensively applied to identify damage in a REB, its meaning is not restricted to it. Other types of fault or abnormal operational conditions, such as unbalanced rotation, misalignments, temperature differences, unsuitable lubricant viscosity, or insufficient supply of it, can significantly change the statistics. The presence of damage in other components, such as gear or pumps, can also be reflected in the REB data. For that reason, the approach chosen here to train the ML model is the envelope spectrum encoded to each feature representing an order band and its value is the adimensionalized amplitude, ranging from 0 to 1, that represents the influence of that particular band in the signal.

The model is conceived to be independent of the amplitude magnitude and the operational speed, thus the data is adimensionalized in both directions, amplitude and frequency. The frequencies are divided by the nominal rotating speed velocity, Ω, in the way the amplitudes, X, are a function of the order, f_{ord}, not the frequency, f. For each signal $X(f)$, the sample process is:

$$f_{ord} = \frac{2\pi f}{\Omega} \tag{58}$$

$$\bar{X}(f_{ord}) = \frac{X(f_{ord}) - \min X(f_{ord})}{\max X(f_{ord})} \tag{59}$$

The encoder takes three parameters: Δ_{fnh} that indicates the feature order frequency span, N is the number of features, and σ, that indicates the percentage of overlap between features. These parameters are optimized along with the estimator hyperparameters. The process consists of taking the adimensionalized ES for each sample, $\bar{X}(f_{ord})$, and reduce it to N number of features. It is accomplished by selecting the maximum adimensional amplitude in a fixed interval defined by Δ_{fnh}. Thus, it results in the final features $S(\not{f})$ to be used to construct the model, each of those being related to a central frequency order $\not{f}[n]$.

$$S[n] = \max \bar{X}[i_n : i_n + \Delta i] \tag{60}$$

$$\not{f}[n] = \frac{f_{ord}[i_n] + f_{ord}[i_n + \Delta i]}{2} \tag{61}$$

where the subscript $n = 1,2, \dots, N$ and i is the index of ES, i_n is the index of the beginning of the interval and it depends on Δi, the number of points within the interval and on the intervals overlap, σ.

$$i_n = i_0 + \ n \cdot \mathrm{round}\big(\Delta i(1 - \sigma)\big) \tag{62}$$

$$\Delta i = \mathrm{round}\left(\frac{\Delta_{fnh}}{f_{ord}[2] \ - f_{ord}[1]}\right) \tag{63}$$

The quality and quantity of information passed on by the encoded features, S, are associated with an appropriate definition of N, Δ_{fnh} and σ values and that should be done specifically for each set of training dataset and model.

There are many methods for hyperparameter tuning as grid search, exhaustive search, random search, or optimization methods. The aim is to test several parameter combinations to choose the one that leads to a better score in accuracy. This work applies Bayesian optimization to select the best set of parameters.

Considering a constant inner-training and validation data set, the pipeline classification outcome can be implemented as only a function of their hyperparameters, having as a target the accuracy error to be minimized. This Black Box problem falls into the category of a multidimensional function, which is expensive to evaluate, with unknown closed forms and gradients. Bayesian optimization is chosen for its time efficiency in this type of problem(BROCHU; CORA; DE FREITAS, 2010). This optimization involves minimizing the errors of predicted results, using an acquisition probabilistic function applied to select the next set of variables to evaluate the error. The acquisition function is approximated by a Gaussian process, and in this case, it selects one of the strategies at each iteration, lower confidence bound, probability of improvement, or expected improvement. This algorithm also balances exploration versus exploitation, explores other ranges of the variables when the variance is high, and exploits the given region, otherwise (HEAD et al., 2020).

As part of the optimization, the parameters that shape the features are defined simultaneously with the estimator and the combination of all parameters results in a unique model. The constraints of the feature parameters guarantee both the algorithm behavior and the physical meaning. The overlap, σ, should be less than 50%, otherwise it would overlap with more than one feature. The Δ_{fnh} must be between the cage order and 1, that is the rotational frequency, to not uphold many peaks and turn the feature dimensions too generalized. The total length of the encoded signal is defined by the number of features, N, which interferes with how many harmonics the code can capture and, consequently, defines the dimension of the data.

Figure 19 shows the same signal encoded with different values of frequency order span Δ_{fnh}. The signal, represented in black, is the acceleration envelope scaled by the maximum amplitude and displayed against the frequency order. The number of features, N, is fixed, and the overlap, σ, is zero. In this way, the smaller the parameter Δ_{fnh}, the greater is the distinction between peaks, while with the same number of features N, higher values of Δ_{fnh} can highlight the highest peaks and cover a wider range of frequency order.

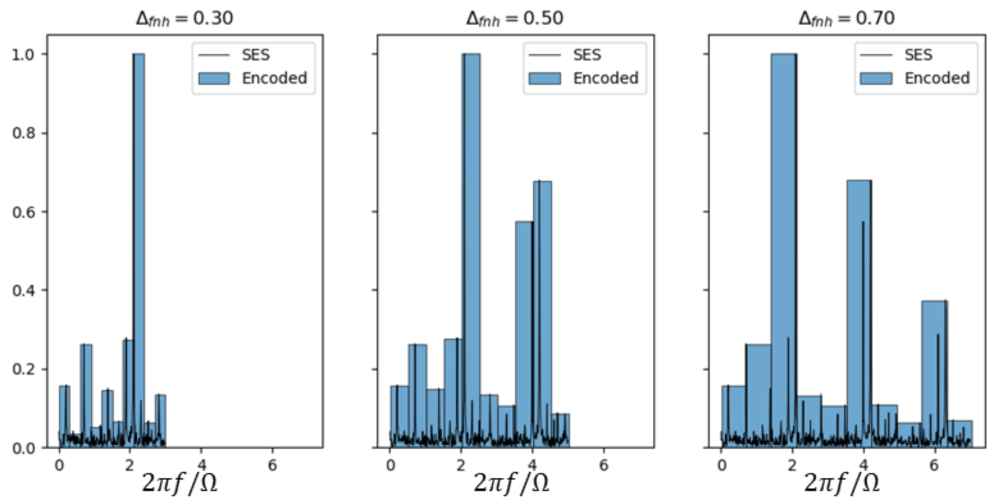


Figure 19. Encoded features with different Δ_{fnh}.

The overlap or distance, σ, from one feature to the next one, is a percentage of the Δ_{fnh},. This application needs to be carefully evaluated since for negative values it distances one feature from the other and can result in a leakage of a significant peak. On the other hand, as close as it gets to 50% positive, the smoother it turns and two different features could bring the same information, as displayed in Figure 20.

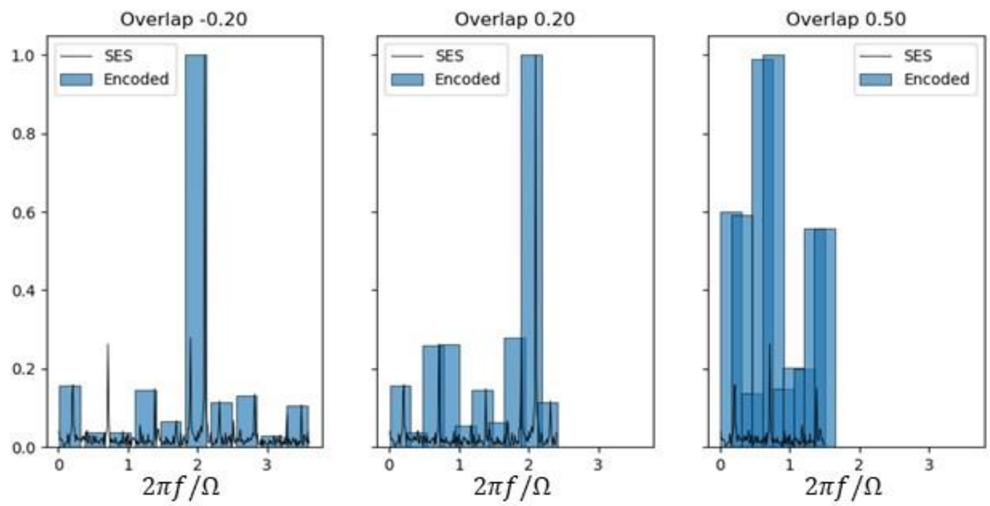


Figure 20. Encoded features variation with overlap.

The last demonstration is on the number of features versus the length of the signal. As illustrated in Figure 20, it must be noted that missing information can lead to imprecise results, but repetitive features can be misleading as well. This parameter should carefully be defined, to not miss any significant peak.

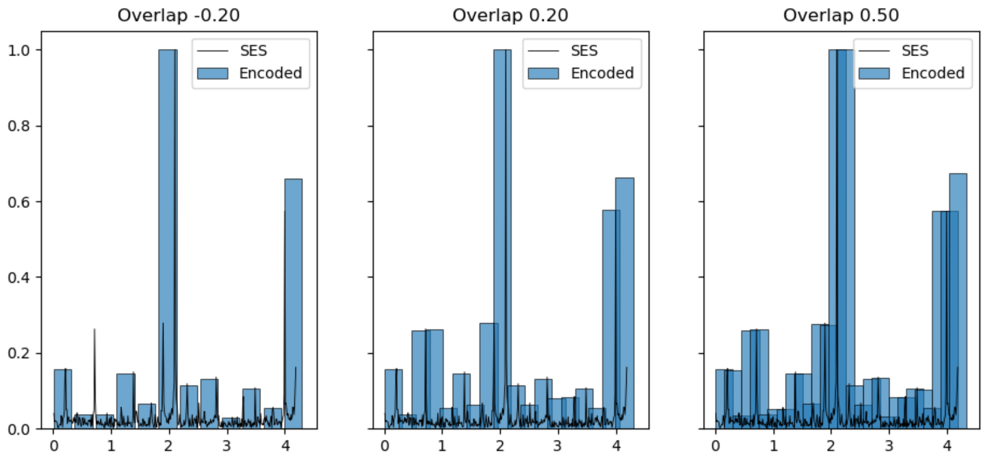


Figure 21. Encoded features with different overlaps and N.

Figure 21 shows the same configuration of Δ_{fnh} and σ, but instead of fixing the number of features N, the frequency order cut off, f_{cut}, were fixed in 4.2 and N was varied according to the relation:

$$N = \frac{f_{cut} - \sigma\Delta_{fnh}}{\Delta_{fnh}(1-\sigma)} \tag{64}$$

The complete procedure is presented in the flowchart in Figure 22 presents the process for fault classification based on a model trained with simulated data and tested with experimental signals unseen by the model until tested. The Measured REB health signals are divided, one bearing data set goes to the Simulated data set, while the other five bearings are part of the real Measured data set. The Parameters search space refers to the SVM hyperparameters, as C, γ and d, and the encoder N, Δ_{fnh} and σ.

The Optimization referenced in Figure 22 is a Bayesian optimization that applies Gaussian processes to decide the next point to evaluate the function (GARDNER *et al.*, 2014). The Black Box is the function to be optimized and its variables are model hyperparameters. The function is called in a predefined N number of times. The target function value is the mean error across all the inner-train-validation K-folds. The number of K-folds is the combination of all the groups in the training set, leaving one out, as the validation set, and training the model with the remainder groups.

Inside the Black Box function, the Pipeline is called to be trained and validated on each fold. The Pipeline object is an algorithm chaining, it sequentially processes the steps of the features encoder, feature scaling, and estimator training or testing. This simplifies the deployment of the ML model and guarantees the validation and test sets pass through the same procedures, with the same parameters, as the model was trained. In this configuration, the encoders and SVM hyperparameters are simultaneously optimized, assuming that the best estimator parameters are not the same for different combinations of features of the same dataset.

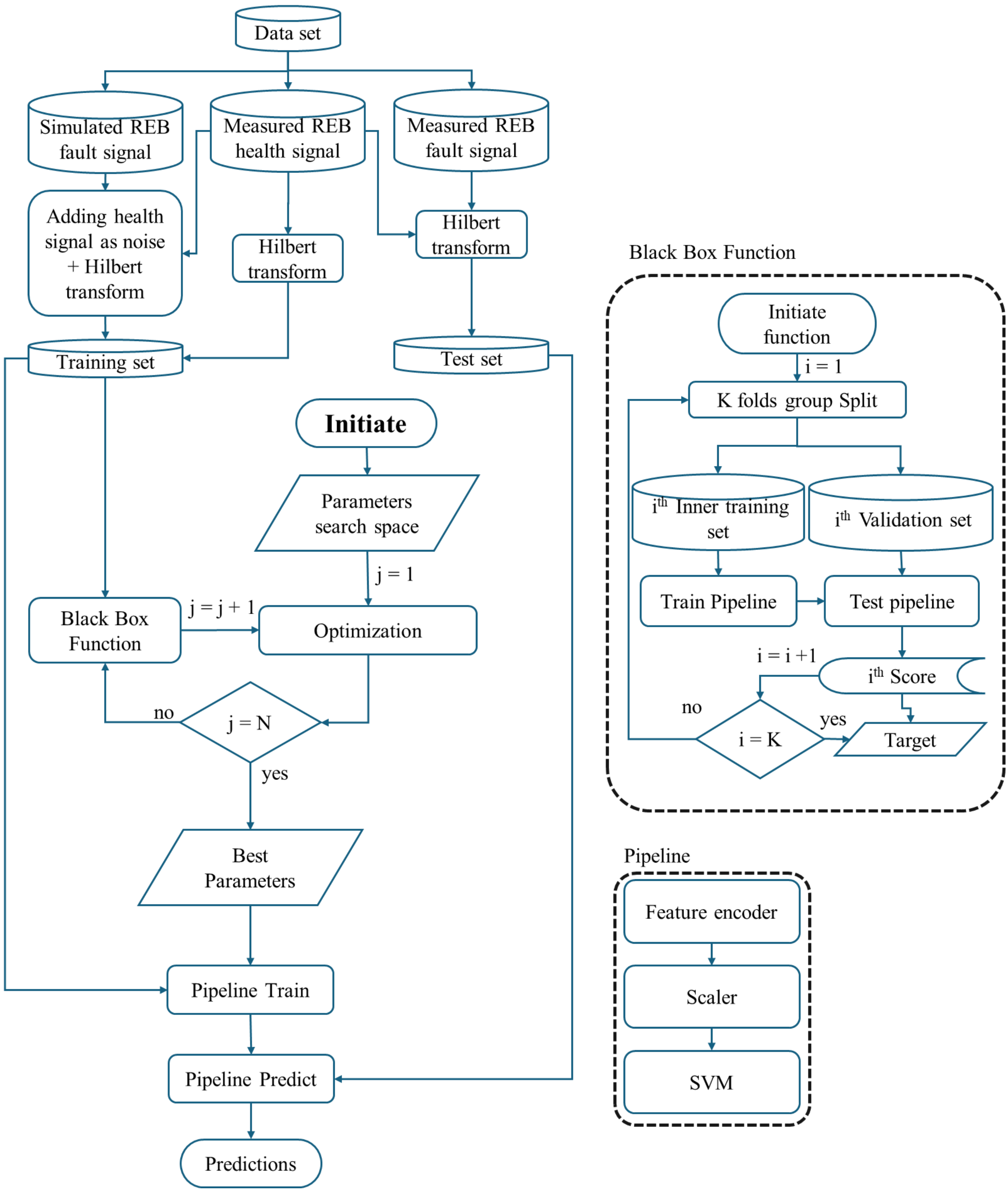


Figure 22. Flowchart for training with simulated faults.

Simulation dataset

To create the data set for this problem, a numerical integration is carried out several times, varying some operational conditions. One way to generate the data would be creating combinations for static load, rotational speed, unbalanced force, and even diametral clearance combined with fault conditions of depth and width.

The bearing-system model can represent the machine dynamic behavior. However, the information in a real signal is quite more complex, and noisier, it may be influenced by the data acquisition instruments or other electrical components, as in the case of the Paderborn dataset. (LESSMEIER *et al.*, 2016)

One possibility is training the model with simulated and real data faults, but this requires the problem of having real data faults correctly labeled. If only real data from health bearings are available, one solution is to mix the signals from the fault simulation with the real bearing response.

It is assumed that the electrical signal does not directly interfere with the system response, only influencing the data acquisition. The concept is to add the health signal as if it is essentially noise and electrical signal contamination. The intention of adding the signals in the frequency domain is to avoid amplitude distortion from different phases and estimate the fraction to be considered in the final signal.

The proposed method takes place before the HT, which is performed in each sample, the process only incorporates the FFT of a real health signal in the same operational condition and the ratio of their root mean square is used to define the balance of both signals amplitudes. The processes are exemplified in the Figure 23 flowchart.

Another randomly generated parameter is the proportion between the RMS of the two signals being mixed. In this way, it is possible to generate different modulation amplitude combinations from the healthy contaminated signal and fault characteristics frequencies. This procedure aims to balance the peaks from both spectrums in a way that one does not overcome the other if there are noise differences in the time domain. It also prevents distortion in amplitudes, as a consequence of phase in the rotational frequency.

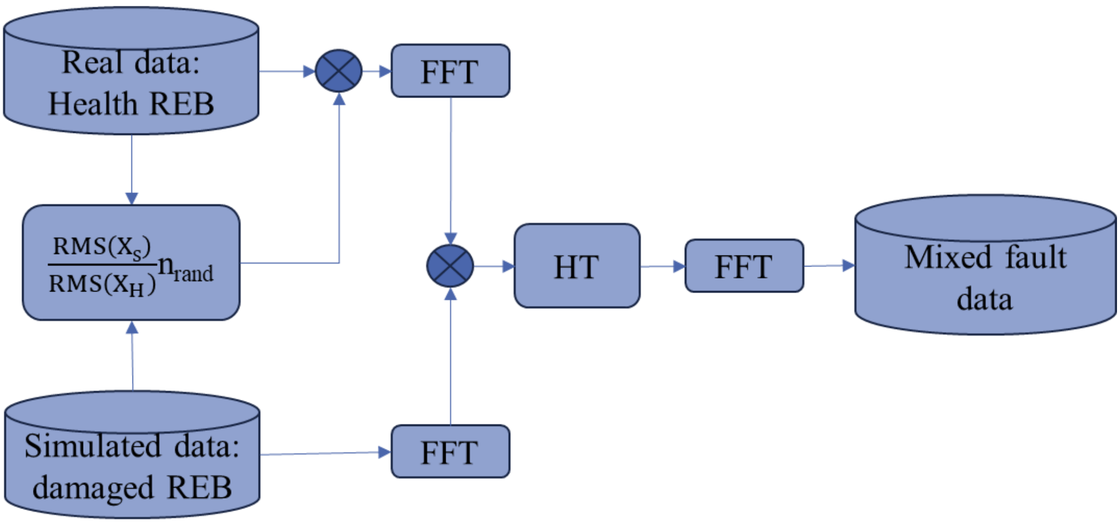


Figure 23. Flowchart of mixing signals.

4 DESCRIPTION OF EXPERIMENTAL DATA

In this chapter validation of the REB fault model is performed using the test bench model from Sawalhi (2007). The REB reaction force, including EHD contact and fault profile described in Chapter 3, is incorporated into the original system equations. The time response is obtained through numerical integration and compared with the original work.

Following the validation of the fault-bearing model, the Paderborn test bench parameters are estimated for a four degrees of freedom system, to include the validated fault REB model. This simulation model is employed to create a dataset to train a machine learning algorithm to identify the health status of measured signals.

Additionally, given the challenges of representing a real-time signal, it is proposed to add the measured healthy data to the simulated fault result, considering only modulation influence from outer sources would be distinguishable in the envelope spectrum.

4.1 Experimental Test Rig: Randall

The validation of the bearing fault identification module demands other simulations and experiment results of a well-known and explored test rig. In this chapter, Sawalhi and Randall (2008) test rig model is employed replacing the existing REB model and parameters by EHD reduced force and localized faults.

The system of equations is based on the Sawalhi PhD thesis and is presented in Appendix A. The test rig scheme shows four REB in a gearbox pedestal. The sensor for data acquisition is on the upper side of the pedestal. The representation of the test rig is presented in Figure 24.

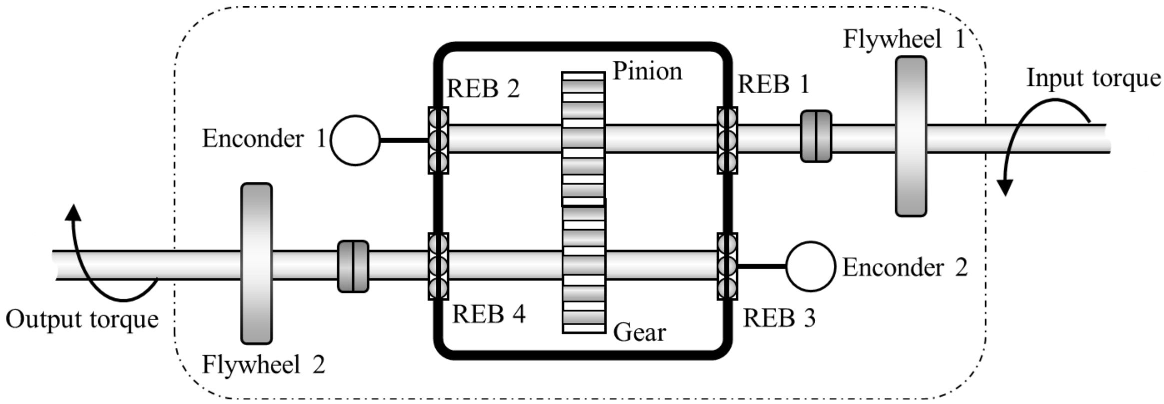


Figure 24. Schematic representation of the Sawalhi and Randall (2008) test rig.

To represent the gear in the system, the stiffness coefficient is interpolated from the data present in the work of Endo (2005), taking into account the variation with the angular displacement and teeth contact. The error and noise are also considered to determine the reaction force connecting both sides of the system.

4.2 Experimental Test Rig: Paderborn

Paderborn University dataset on ball bearings is one of the most extensive collections of vibration data designed for fault identification. The collection consists of data from 32 groups, including 6 undamaged bearings and 26 bearings with internal surface damages, divided into 12 artificially generated and 14 with accelerated-life-induced faults. Each group corresponds to a specific bearing, according to Table 1

Table 1 Paderborn Dataset REB description.

Prefix	Status	Accelerated life time	Artificially damaged bearings
K0	Health	K001, K002, K003, K004, K005, and K006	
KA	Outer race fault	KA04, KA15, KA16, KA22, and KA30	KA01, KA03, KA05, KA06, KA07, KA08, and KA09
KI	Inner race fault	KI04, KI14, KI16, KI17, KI18, and KI21	KI01, KI03, KI05, KI07, and KI08
KB	Inner and outer race fault	KB23, KB24, KB27	

Each group of bearings contains 4 distinct operation conditions with 20 samples each, all of them running for 4 seconds of vibration data, that is, every bearing group have 80 samples but only with 4 variations. The operational conditions are presented in Table 2.

Table 2 Data set operational conditions.

Condition	Rotation [RPM]	Static load [N]	Torque load [Nm]
0	1500	1000	0.7
1	900	1000	0.7
2	1500	1000	0.1
3	1500	400	0.7

4.3 Numerical Simulation: Data Generator

The Paderborn University laboratory Kat (*Konstruktions- und Antriebstechnik*) test bench model used here is designed to be representative of the main dynamic behaviors without lead to excessive computational time, for that, only 4 degrees of freedom, 2 for the shaft and 2 for the REB inside housing are considered. The information used to build the model are given in the images and descriptions in Lessmeier et al. (2014). The model for the bearings reaction force added to the rotor system considers the following assumptions:

- The inner race rotates at the shaft constant speed.
- RE always remain in contact with the races, excepting in case of passing by a spall deeper than the displacement.
- All translational motions are in-plane
- The balls and cage are assumed with negligible masses,
- The motion occurs in y-x plane and rotations are about z-axis.

Figure 25 shows the model for the inner housing set up. The coupling between shaft and inner housing is given by the reduced order model for the bearing reaction force, f_{bx} and f_{by}, given by Equations (46) and (47), as a function of each element j displacement, δ_j, velocity, $\dot{\delta}_j$ and angular position φ_j, given by Equations (17), (20) and (19) respectively.

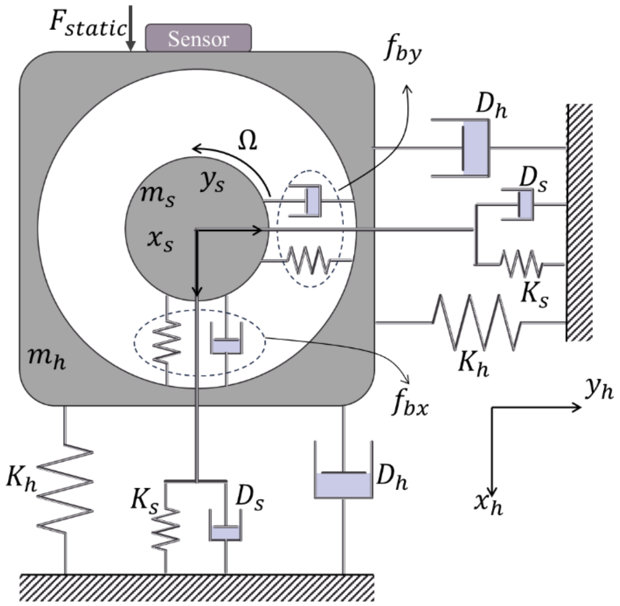


Figure 25. REB test bench model for Paderborn Kat dataset.

The inner housing has mass m_h, stiffness coefficient, K_h and damping coefficient D_h . The sensor is located in the housing, so the signal is in the x_h direction, the same as the experimental test rig. The shaft has a lumped mass of m_s and the bearing is represented as a system of springs and dampers located in the balls contact.

The parameters K_s and D_s are the shaft stiffness and damping coefficients, the shaft passes across the REB inner ring, being isolated by the couplings and self-aligning REB. These forces restrict the inner part of the bearing, so it does not run loose, once the static force is applied to the outer ring.

$\boldsymbol{\delta}_R$ is the radial displacement of the center of the shaft relative to the outer ring, fixed in the housing, as described in Lessmeier et al. (2014). The components of the displacement and velocity in *x* and *y* direction are applied in the Equations (17) and (20) and given by :

$$\delta_{rx} = x_s - x_h \text{ and } \delta_{ry} = y_s - y_h \quad (65)$$

$$\dot{\delta}_{rx} = \dot{x}_s - \dot{x}_h \text{ and } \dot{\delta}_{ry} = \dot{y}_s - \dot{y}_h \quad (66)$$

The complete system of equation has four degrees of freedom is given by:

$$m_h\ddot{x}_h + D_h\dot{x}_h + K_h x_h = f_{bx}(x_s, x_h, \dot{x}_s, \dot{x}_h, t) + F_{static} \quad (67)$$

$$m_h\ddot{y}_h + D_h\dot{y}_h + K_h y_h = f_{by}(x_s, x_h, \dot{x}_s, \dot{x}_h, t)$$

$$m_s\ddot{x}_s + D_s\dot{x}_s + K_s x_s = -f_{bx}(x_s, x_h, \dot{x}_s, \dot{x}_h, t) + F_{me}\cos\Omega t$$

$$m_s\ddot{y}_s + D_s\dot{y}_s + K_s y_s = -f_{by}(x_s, x_h, \dot{x}_s, \dot{x}_h, t) + F_{me}\sin\Omega t$$

the unbalanced force, F_{me}, that may occur simultaneously with the bearing, has the modulus $F_{me} = m_e\varepsilon\Omega^2$, being the residual mass, m_e, and eccentricity radius of ε, transmitted to the center of the shaft and Ω, the rotational speed. The static force, F_{static}, is applied to the outer race throughout a screw tightened between the inner and outer housing box.

The parameters to model these products were determined based on the information given by the original work and data extracted from the time signals. The definition of the bearing parameters is based on the geometrical and operation conditions, considering the lubricated EHD regime. The rolling bearing is not completely isolated from other components and the sensor is located on the housing, hence the vibration is assessed throughout the inner housing, so its influence has to be taken into consideration. The list of components and the assumptions on their influence on the signal follows:

- Shaft – connecting the inner race of the test bearing to the self-aligned ball bearings, with stiffness K_s, transmitting only a residual unbalanced force.
- Internal housing, has mass m_h and connects the outer ring of the test bearing to the outer housing with stiffness coefficient K_h
- External housing is considered fixed.
- Self-aligned ball bearings are considered fixed.
- Spherical bearing is considered part of the inner housing.
- Internal oil supply: the test bearing is immersed in oil retained by mechanical seals inside the inner housing. It is represented by a damping coefficient D_s.
- Test bearing connects the shaft to the inner housing through the nonlinear reaction force.
- Frequency inverter has its signal added to the simulations as background noise.

The main idea is to identify the most influential parts of the mechanical system to implement a simple mathematical model. A lumped parameters system with only two degrees of freedom, representing the vertical direction, is applied to identify the most influential parameters:

$$\begin{bmatrix} m_h & 0 \\ 0 & m_s \end{bmatrix}\begin{Bmatrix} \ddot{x}_h \\ \ddot{x}_s \end{Bmatrix} + \begin{bmatrix} D_h + D_{bl} & -D_{bl} \\ -D_{bl} & D_s + D_{bl} \end{bmatrix}\begin{Bmatrix} \dot{x}_h \\ \dot{x}_s \end{Bmatrix} + \begin{bmatrix} K_h + K_{bl} & -K_{bl} \\ -K_{bl} & K_s + K_{bl} \end{bmatrix}\begin{Bmatrix} x_h \\ x_s \end{Bmatrix} = \begin{Bmatrix} F_{static} \\ F_{me}\cos\Omega t \end{Bmatrix} \quad (68)$$

K_{bl} is the linearized bearing stiffness, only used to estimate the system critical frequencies:

$$\mathbf{Q}_{static} = K_{bl}\boldsymbol{\delta}_R \quad (69)$$

$$K_{bl} = Q_{static}\left(\frac{K\Gamma_d}{Q_{static} - \Delta F\Gamma}\right)^{\frac{1}{d}} \quad (70)$$

given that $Q_j = Q_{static}\cos\psi_j$, is the projection of the load on the radial direction to each rolling element. Then, Γ and Γ_{d} are a consequence of the $\cos\varphi_j$ summation in the load zone. If the diametral clearance, C_r, is null, then only the elements with the azimuth position $\varphi_j \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, should be considered (HARRIS, 1991).

$$\Gamma = \sum_{j=1}^{Z} \varrho_j \cos\varphi_j \quad (71)$$

$$\Gamma_d = \sum_{j=1}^{Z} \varrho_j \cos^{d+1}\varphi_j \quad (72)$$

the binary variable ϱ_j is zero if φ_j lays outside the load zone, and ϱ_j is one, otherwise.

For this initial study, only the hertz model represents a dry contact, to first establish the load acting in the bearing. Shaft mass, m_s, and internal housing mass, m_h, are estimated based on the inner ring dimension and the drawing present in Lessmeier et al. (2014).

The system data with random faults were used as an effort to obtain a broad-spectrum excitation, allowing to visualize the critical frequencies of the system. The system parameters were defined based on the PSD (Power spectrum density) and Fourier transform. The approximated locations of the critical frequencies were identified as f_{n1} and f_{n2}. The approximation for K_h and K_s is:

$$\frac{K_h + K_{bl}}{m_h} \cong \left(2\pi f_1\right)^2 \rightarrow \; K_h \cong m_h\left(2\pi f_{n1}\right)^2 - K_{bl} \tag{73}$$

$$\frac{K_s + K_{bl}}{m_s} \cong \left(2\pi f_{n2}\right)^2 \rightarrow \; K_s \cong m_s\left(2\pi f_{n2}\right)^2 - K_{bl} \tag{74}$$

Although the real values of K_s and K_h cannot be determined, Equations (73) and (74) reduce the system to two uncoupled spring-mass system, which can provide an estimative for these parameters magnitude. There are numerous of data on these tests, and it is not realistic to use them all since an amount of suitable data is not usually available. So only the vibration signal of a healthy bearing is used to build the test bench model.

5 RESULTS AND DISCUSSION

5.1 Bearing model validation

As previously stated, the validation of the bearing model is done using the bench model from Sawalhi, 2008, and the bearing reaction force, with EHD contact and fault profile described in Chapter 3, is inserted into the system equations. The time response is obtained through numerical integration.

Figure 26 considers an EHD lubrication for a rotation speed of 600 rpm and load of 560 N for a two-row auto-compensating ball bearing with pitch diameter of 38.5 mm and 7.12 mm ball diameter. The parameters for the reduced order force are obtained by the EHD simulations described in the section 3.2. The equivalent stiffness coefficient is $K = 6.02 \times 10^9 \mathrm{N/m}^d$, the exponent of the RE displacement, $d = 1.48$, the residual force, $\Delta F = 1.83\ N$ and damping coefficient is $D = 16.13\ \mathrm{N.s}$.

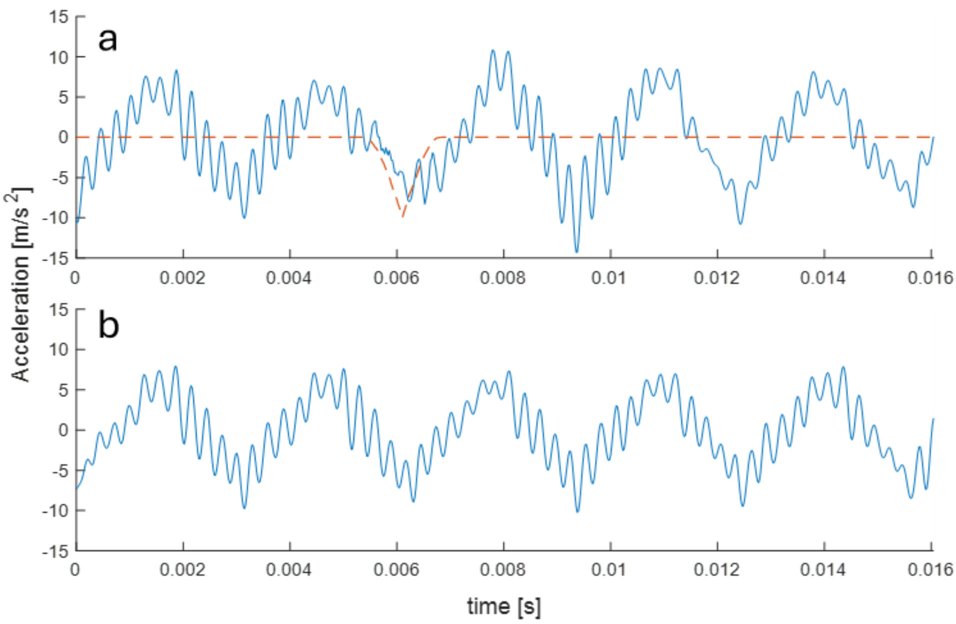


Figure 26. Time response (a) outer race fault and (b) health REB.

The main vibration source in the acceleration signal comes from the gearbox, modulated by the 32-tooth period. Figure 26 (a) shows the fault bearing signal and the exact moment the ball passes through the fault, by displaying the smooth β profile by the red-dashed-line. As β ranges from 0 to 1, the example is scaled from 10:1. The spall has a 2º angular span, which is 0.8 mm width, with a maximum depth of 19 μm. Comparing both signals, with and without the fault, in Figure 26 (a) and (b), the fault does not significantly change the time signal, which is dominated by the gear modulation.

Figure 27 (a) and (b) shows respectively the Power Spectral Density (PSD) for the simulation and measured signals from (SAWALHI, N.; RANDALL, 2008b) and can be compared with the PSD in Figure 27 (c), illustrating the EHD simulated signals. I was applied the Welch method, 50% overlap, with 512 points and Hanning windowing. The signal used is the acceleration of the sprung mass on the OR fault bearing casing, representing the sensor reading point, added by pink noise with signal to noise ratio (SNR) of 20 dB.

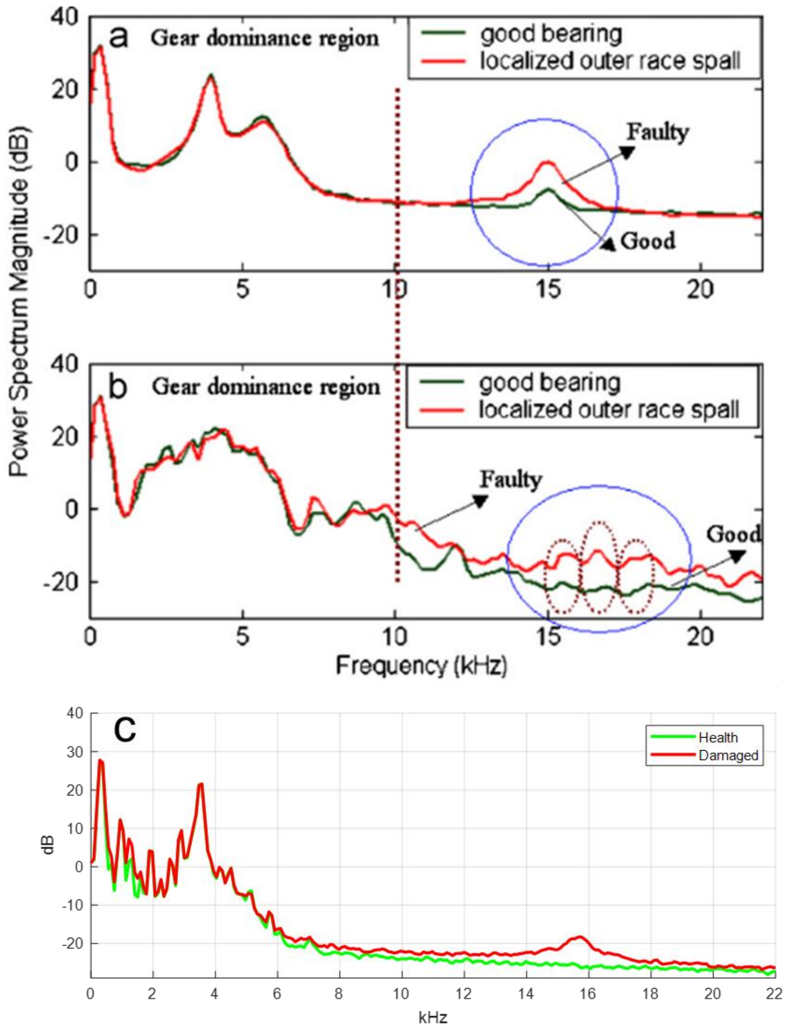


Figure 27. PSD comparing Health and Damaged signal of REB : (a) simulated, (b) experimental from (SAWALHI, N.; RANDALL, 2008b) *and (c) simulated with the EHD simulations and smooth fault function.*

The first peak at 320 Hz is the teeth frequency for the rotational frequency of 10 Hz. Therefore, the gear response dominates the PSD below 6 kHz, and the major difference due to the fault can be seen around the spring-mass critical frequency, near 15kHz. This is similar to a transducer natural frequency on the HFFT analysis. The signal is obtained by the variable of this degree of freedom and its high natural frequency region is excited by the fault.

To access the envelope spectrum with the most information about the OR fault, the fast kurtogram is applied to identify the carrier frequency and frequency resolution with higher values of kurtosis, indicating the presence of sharp peaks in the signal. This is characteristic of early-stage faults.

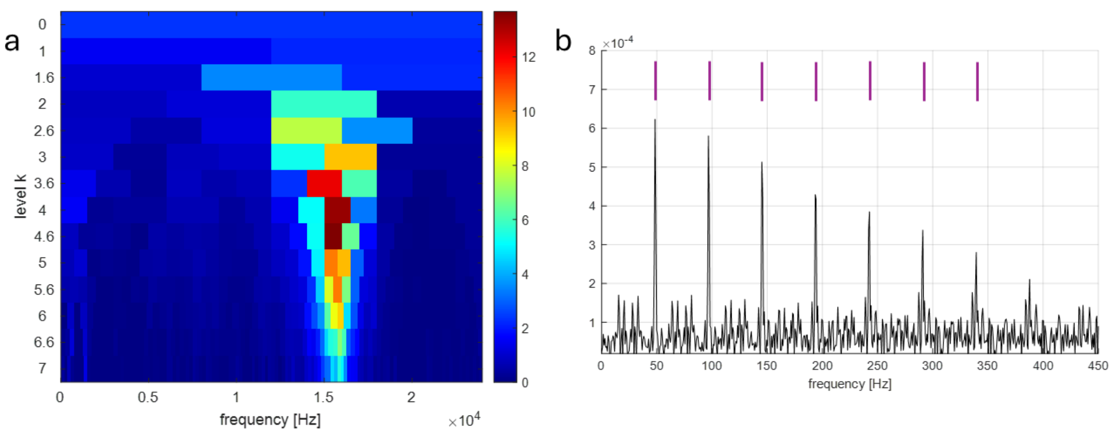


Figure 28. Outer race fault (a) Fast Kurtogram and (b) filtered envelope.

The optimum carrier frequency is similar to the one in the PSD, 15.75 kHz, and the signal is filtered in a resolution of 1 kHz. The peaks in the filtered envelope spectrum, seen in Figure 28 (b), are spaced by the BPFO frequency of 49Hz, and its multiple frequencies. This indicates the impact on the OR fault, excited the spring mass, causing a transient that enables the identification, namely, the gear modulation is not as important in higher frequencies.

The fault in the inner race has the same length, 0.8 mm, with a maximum depth of 27 μm. Figure 29 shows the time signal with (a) and without (b) fault on the inner race. As this model contemplates the slippage between the inner race and the cage, it results in a relative velocity between the RE and the inner race, oscillating around the nominal ω_c. This effect can be observed in the forced path dashed line, in Figure 29 (a), when exiting the fault, the RE oscillates before returning to the normal path. The slippage effect is also considered in the outer race fault, however as the fault spall is stationary, it does not affect the β function.

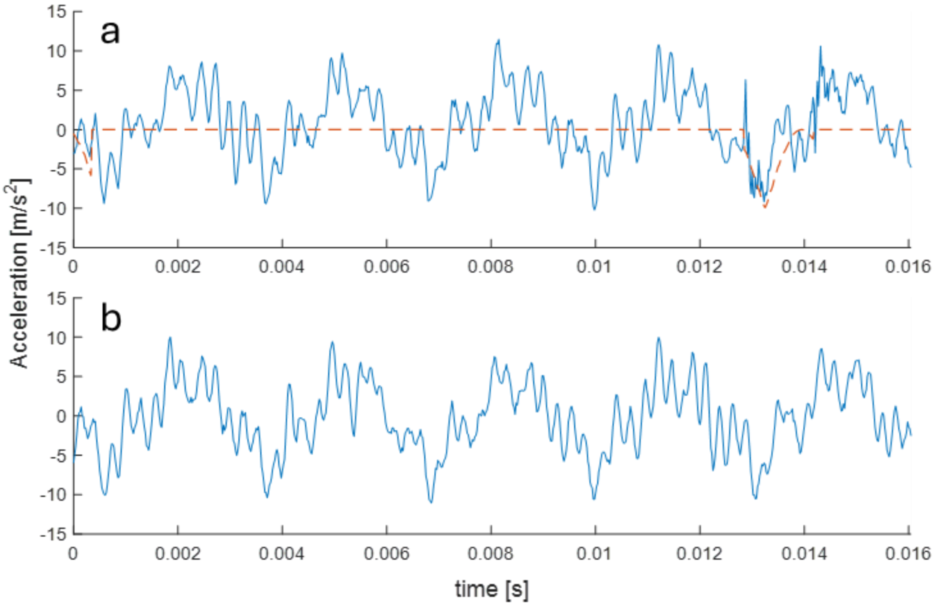


Figure 29. Time response (a) inner race fault and (b) health REB.

Figure 30 (c) suggests that the most effective frequency band to filter the signal, is the one centered at 15 kHz with a bandwidth of 6 kHz. In Figure 30 (d), the envelope spectrum shows the BPFI of 71 Hz, with side bands of the 10 Hz, the rotational frequency. This can be compared with the FK and filtered envelope spectrum from the experimental data presented in (SAWALHI, N.; RANDALL, 2008b).

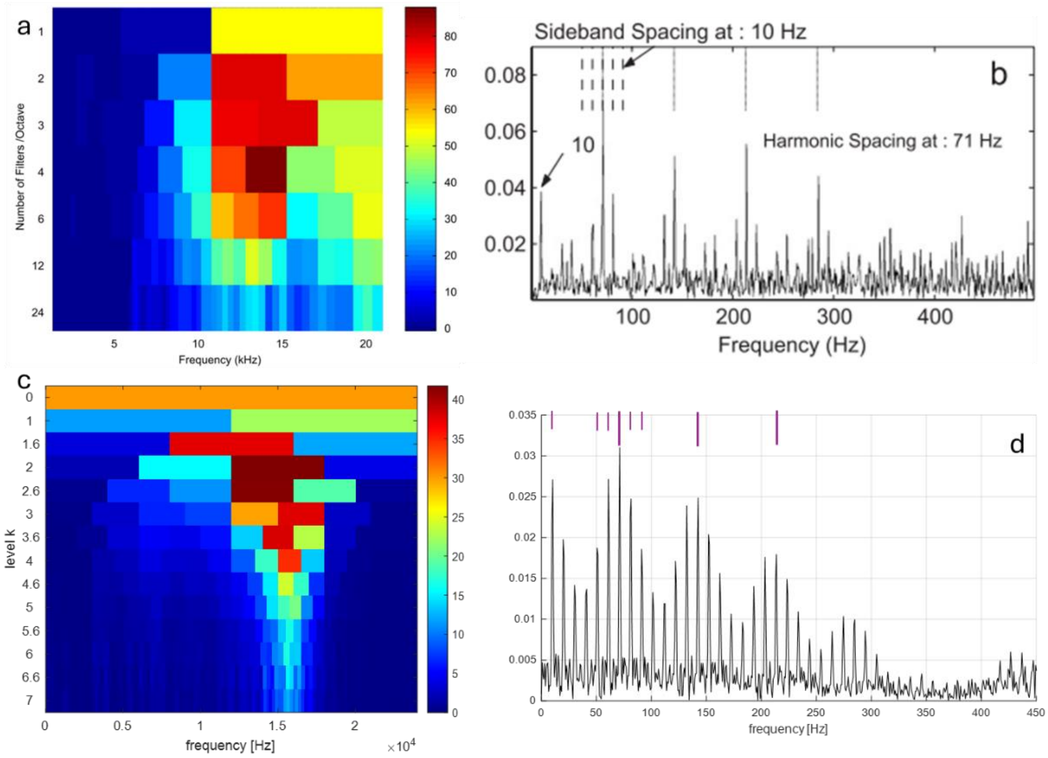


Figure 30. Inner race fault (a) Fast Kurtogram and (b) filtered envelope for the experimental measured signal from (SAWALHI, N.; RANDALL, 2008b).*(c) Fast Kurtogram and (b) filtered envelope for the EHD simulation.*

The fault in the rolling element was also simulated and the FK and envelope are presented in Figure 32 (a) and (b), respectively. The fault is an engraved rectangle, with 3.6 mm width and 0.5 mm depth. The included angles in the spherical element are represented by Figure 31 , 60° and 8°, so the last will determine the angular span of the fault in contact with the inner race, 1.8º, and a depth of 7.3 μm, and the angular span with the outer race, 1.2°, with a depth of 10.5 μm.

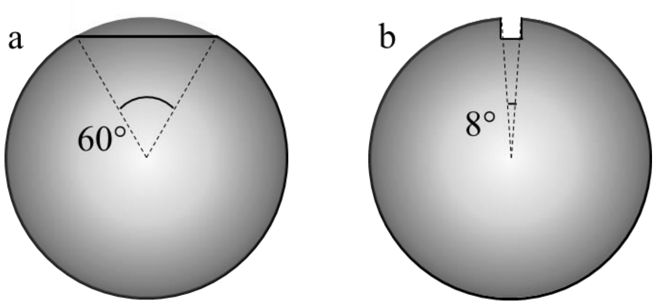


Figure 31. Spall in the spherical rolling element with 60° inclusion angle in view a and 8° inclusion angle in view b.

The spin frequency of 26 Hz and its multiples appear in the envelope spectrum with 4 Hz side bands, cage rotational frequency, which also is the first amplitude in the envelope spectrum. The 2xBPF which is 52 Hz, appears with similar amplitude since the RE hits both races one each complete spin.

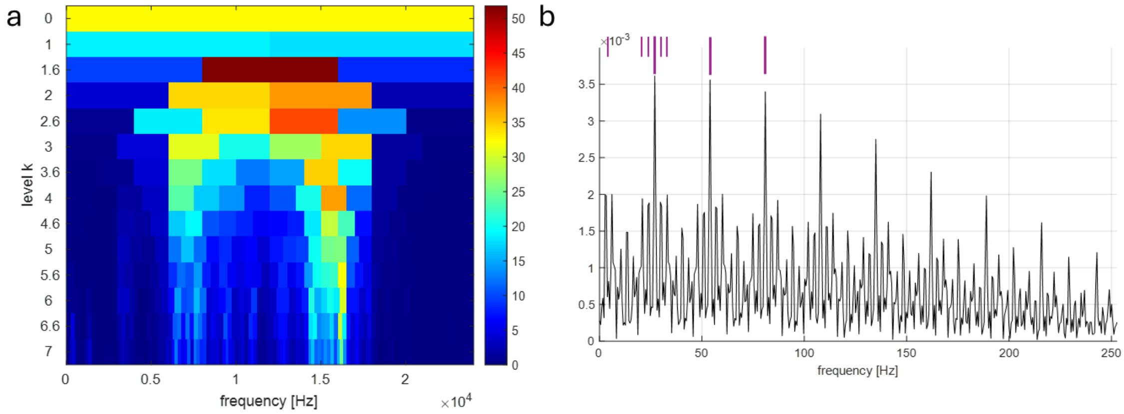


Figure 32. Ball fault (a) Fast Kurtogram and (b) filtered envelope.

5.2 Paderborn Model

As the published works on the Paderborn test bench, did not include a numerical model, or any model analyses or parameters, one has to be made to incorporate the model of the REB validated in the previous section. The simulated Paderborn model parameters, time, and frequency response will be presented, as well as a detailed examination of the internal REB contacts passing through a fault.

The construction of the Paderborn test bench model is based on the information provided by (LESSMEIER et al.,2014; 2016) and the data on the vibration signals. Two health REB signals were investigated to estimate the modal parameters of this system. The PSDs for bearings K001 and K003 are displayed in Figure 33. Those are composed of two noisy signals with the envelope spectrum displaying only the current electrical frequencies. The white noise can excite a broad band in the system and allows the identification of the main characteristic frequencies of the system.

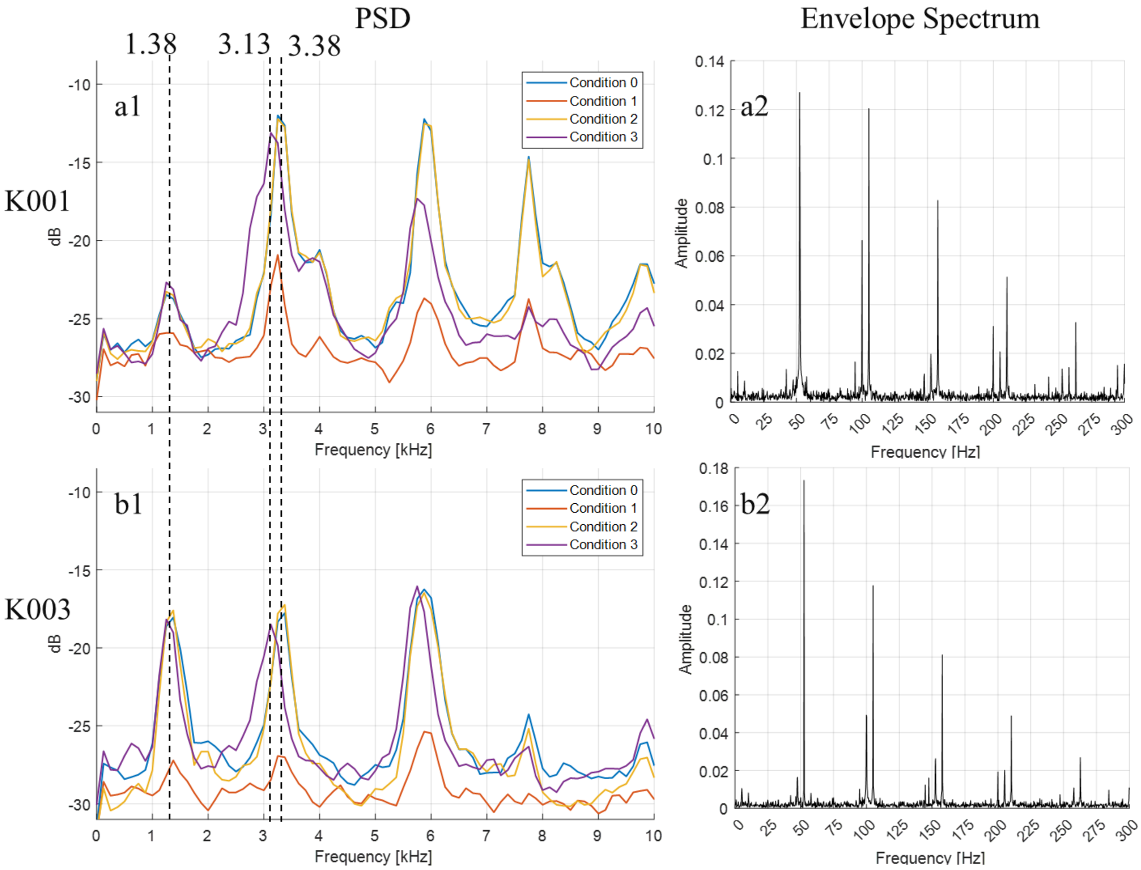


Figure 33. a1) PSD for 20 samples of all four conditions of the REB K001 and a2) is the envelope spectrum for condition 0; b1) is the PSD for 20 samples of all four conditions of the REB K003 and b2) is the envelope spectrum for condition 0.

The system is intended to have only two DOFs in each direction, so the frequencies of 1.38 kHz and 3.3 kHz were selected. The behavior around them is similar in both REB, where condition 3 has a slightly lower frequency, about 7%, as the system is subjected to a lighter load, and condition 1 has the smallest amplitudes, as the excitation from the residual unbalanced mass is proportional to the squared rotational frequency, and condition 0 and 2 are practically the same.

Applying Equation (70) with to REB dry contact parameters, the linearized stiffness is $K_{bl} = 4.25 \times 10^7$ N/m. The mass in the bearing node, m_s, is 0.32 kg and the mass in the housing node is, $m_h = 3.2$ kg. These estimates are based on the REB 6203 dimensions and the inner housing section view presented in the schematic figure (LESSMEIER et al., 2014).

The stiffness between the outer housing and the bearing external ring is approximated to $K_h = 1.9 \times 10^8$ N/m. The first frequency estimate of the inner housing is given by Equation

(73), and the second frequency employed in Equation (74), gives an estimative of the shaft stiffness of $K_s = 1 \times 10^8$ N/m.

The linearized frequency response, Figure 34, provides a model to adjust the damping coefficients, $D_h = 2500$ N.s/m and $D_s = 800$ N.s/m. This is the base to determine the modal parameters of the system. The first critical frequency is 1314 Hz, with a damping factor of 4.9%, and the second is 3490 Hz with 5.7% damping factor.

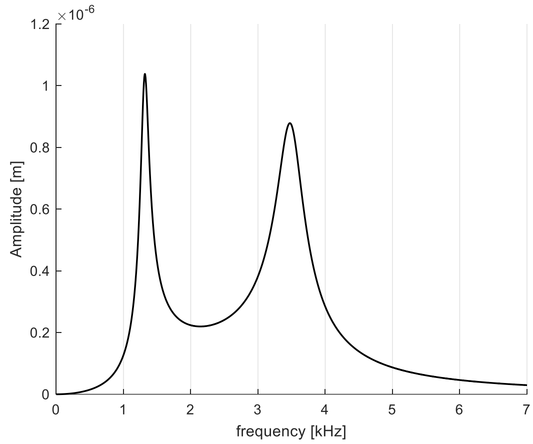


Figure 34. Linearized frequency response due to unbalance in the bearing node.

With the system parameters defined, the operation conditions of rotational speed and static load are used to characterize the reduced force parameters. Figure 35 (a) shows the reaction force for conditions 0 and 2, 1500 RPM and 94 N, (b) for condition 900 RPM with 94 N, and (c) for condition 1500 RPM with 30 N.

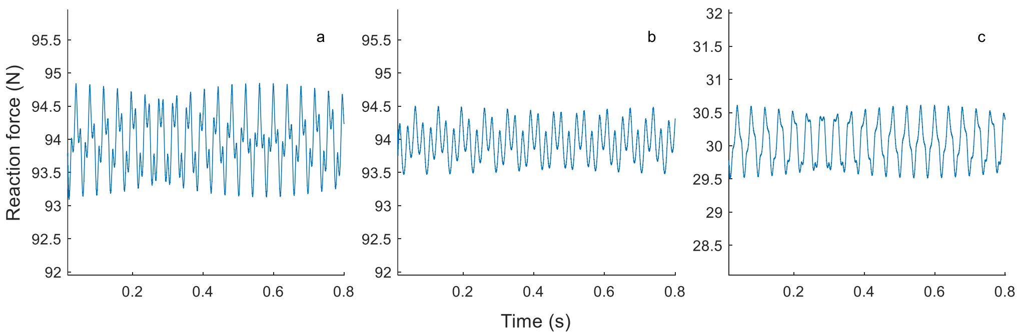


Figure 35. Simulated Reaction force in the bearing center with zero clearance in all conditions a) conditions 0 and 2, b) condition 1, and c) condition 3.

In Table 3, the reduced force parameters are presented for the data set operation conditions. Conditions 0 and 2 only differ in torque, which is not a variable in this model, being regarded as the same.

Table 3 EHD reduced model force parameters.

Conditions	K	d	ΔF	D
0 and 2	5.46×10^9 N/m$^{1.46}$	1.46	2.90 N	22.0 N.s/m
1	6.20×10^9 N/m$^{1.48}$	1.48	2.11 N	14.2 N.s/m
3	3.14×10^9 N/m$^{1.42}$	1.42	2.41 N	26.3 N.s/m

Figure 36 shows the details of the relative displacement and velocity of one RE throughout a spall passage. The forced paths are represented by the dashed lines, imposed by the fault model. In both cases, negative values of displacement and velocity are not accepted and are assumed to be null.

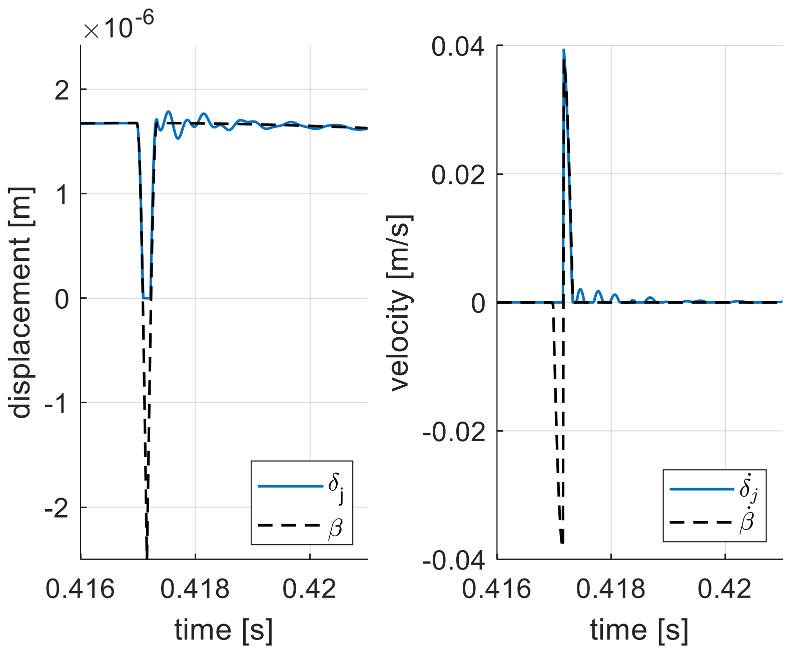


Figure 36. Displacement and velocity of the RE passing through a spall.

For that reason, the intensity in which the RE is drawn to the fault is a function of the maximum depth that must be constrained by the fault geometry, or even it can be set to less if the fault is shallow. This factor is one of the uncertainties used to create a diversified dataset.

Figure 37 presents two severity cases in response to an outer race fault. Both cases are accomplished on the operation condition 0, namely, 25 Hz and 1 kN in the inner housing, null clearance, and an unbalance of 10^{-5} kg.m². There are two fault severities: cases (a), (c), (e) and (g) regard a fault angular span of 9° and depth of 25 µm, only 10% of its maximum constraint. Meanwhile, in Figure 37 (b), (d), (f) and (h), the spall is only 1° angular span and 2.9 µm deep, which in this case is the maximum depth constrain. The lines of f_{1x}, f_{2x} and f_{3x} represent the x-component of the reaction force of the three consecutive RE always passing through the spall in the static force direction, added up to the reaction force in the REB center, f_x. The sinusoidal wave pattern of the REs displacement and force is an outcome of the orbital projection into the vertical direction. If the RE is unloaded the displacement is null, which can be seen in the unloaded zone and deep spalls.

The forced displacement imposed by the spall causes momentaneous unloading in the RE when passing through it. Figure 37 (a) clearly shows the RE complete loss of contact with the inner race. Consequently, this RE contact is not compressed anymore, being necessary that the adjacent REs take the extra load, increasing their displacements and reaction forces, although not enough to avoid its effect on the center of REB, as Figure 37 (b). As the fault area is narrow enough the RE passage can be seen as an impulse.

The difference in the spectrum is a well-known occurrence, namely, the incipient faults excite higher frequencies, while extensive faults have a spectrum with large amplitudes in lower frequencies (RANDALL; ANTONI, 2011)

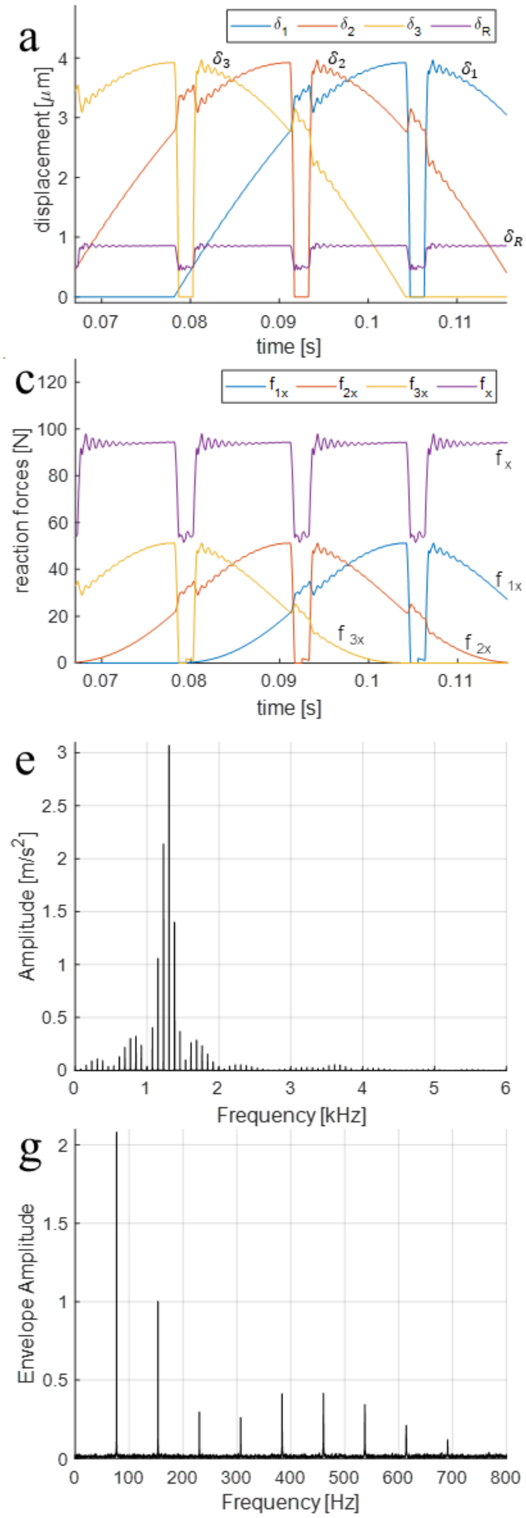


Figure 37. Bearing response of outer race fault in condition-0 under two cases of severity, (a) and (b) are displacements of three RE and the central node of the bearing; (c) and (d) are the reaction forces; (e)-(h) refers to the inner housing signal, being (e) and (f) the acceleration spectrum and (g) and (h) the envelope spectrum.

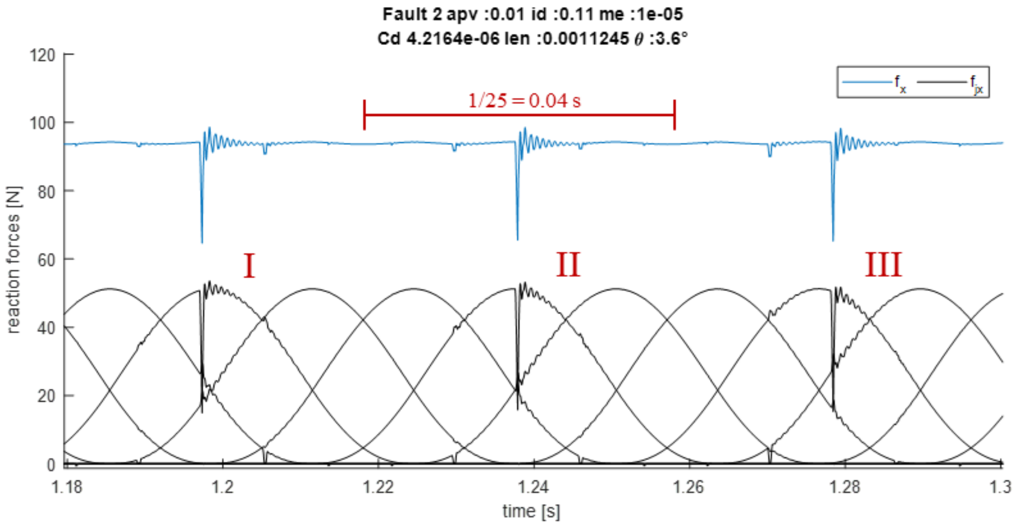


Figure 38. Reaction force of the REB center (f_x) and the reaction forces of the RE (f_{jx}) in the x-direction in the occurrence of inner race fault.

Figure 38 signal is modulated by the rotational frequency. There is a difference between positions and encounters of the RE and inner race. Therefore**,** each time the damage passes through the load zone the impact happens in a slightly different position, represented by regions I, II and III. The complete turn of the inner race takes 0.04s for a 25Hz rotational frequency, and the peak has just one major transient.

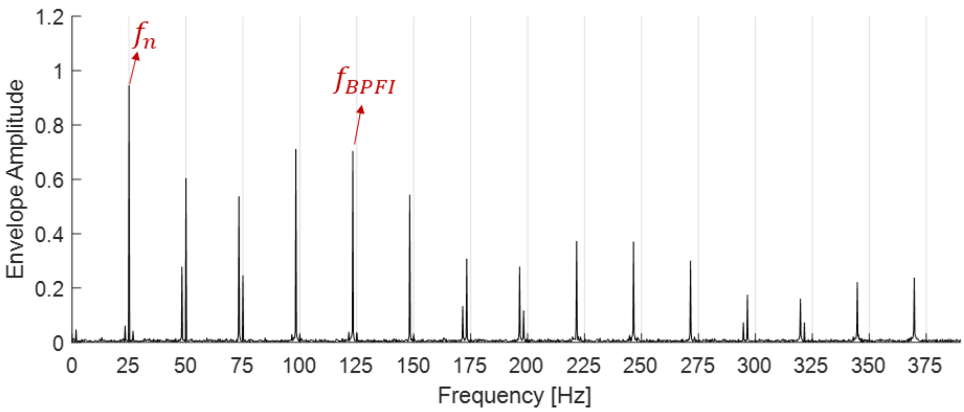


Figure 39. Envelope spectrum of the vibration signal of inner race fault simulation.

The envelope spectrum of the case displayed in Figure 38 is presented in Figure 39, where the first and highest peak is the rotational frequency of 25 Hz, and its harmonics are combined with the BPFI of 123 Hz.

5.3 Paderborn Data set

In this section, a brief signal analysis of the measured vibration from Paderborn data set is carried out. The intention is to clarify the influence of external signals and evaluate if methods such as fast kurtogram are of use for these specific cases.

The main challenge in the identification of these bearings is the electrical current. The inverter with 16 kHz switching frequency appears strongly on the undamaged bearings. This signal is modulated by the current frequency and added to the bearing signal. These amplitudes do not cause visible transients and can get in the way of identifying frequencies of interest. This can be the case when using automated tools to detect and filter high kurtosis regions to improve the amplitudes for envelope analysis. In a close look, the sharp shape of impulses of early-stage damages and the inverter can be similar.

Analyzing the envelope spectrum from the undamaged bearings, it is clear the modulating frequencies, unrelated to the rotational speeds take over the response. As the samples of the envelope spectrum for 900 rpm rotation, the peaks on 52.5Hz and 100 Hz with sidebands of 5.25 are similar to those on the samples running at 1500 rpm. This is influenced by another source independent of the bearing rotational elements, probably from the electrical influence. As mentioned by the authors (LESSMEIER *et al.*, 2016), frequency inverter was deliberately left unfiltered to resemble an industrial application challenge.

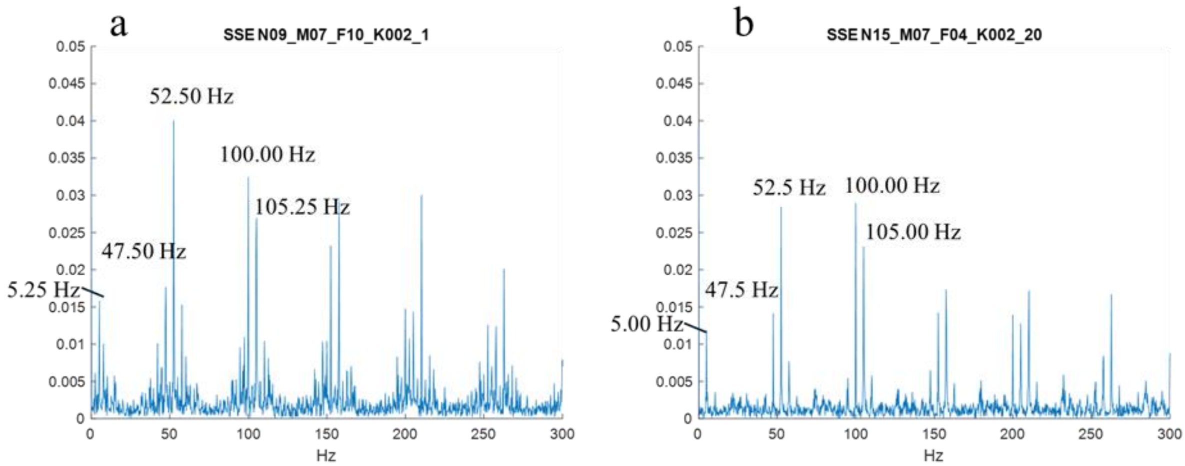


Figure 40. Envelope spectrum of samples (a) N09_M07_F10_K002_1 e (b) N15_M07_F04_K002_20.

The frequency inverter has a switching frequency of 16 kHz, and it is modulated by the electric current frequency of 50 Hz. Therefore, harmonics of the current frequencies appear at the beginning of the envelope spectrum, as can be seen in Figure 40.

The amplitude of modulation in the health bearing can be lower than in the BPFO and BPFI signals, depending on the experiment conditions and the type and extension of the damage. Filtering the signal by the frequency band with higher kurtosis, as in section 4.1, to find the optimum resolution to identify the fault characteristic frequencies can result in selecting the regions where the current frequency is dominant.

Figure 41 shows the Fast Kurtogram of sample 1, case 0, bearing KA04. The frequency band with higher values of kurtosis is the carrier frequency 26.7 kHz, with a bandpass of 10.7 kHz. The resultant filtered envelope shows the electrical frequency of 51.25 Hz and its multiples.

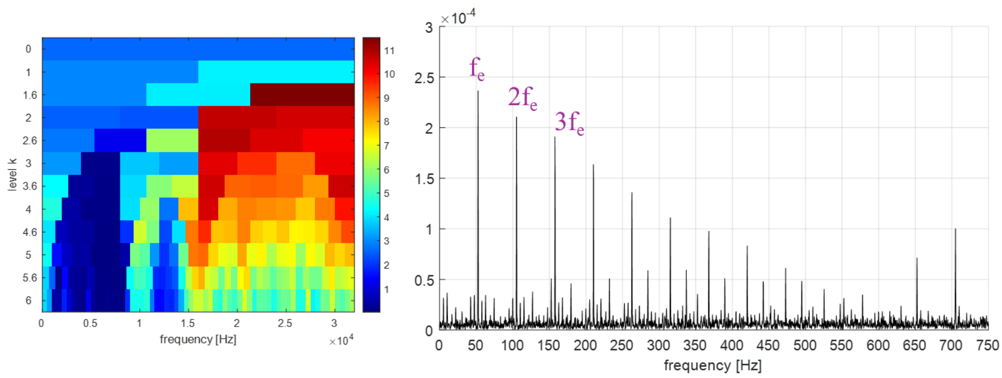


Figure 41. FK and envelope spectrum filtered in the indicated frequency band for sample N15_M07_F10_KA04_1.

While just applying the Hilbert transform without filtering, for the same sample, the result is the BPFO of 76.25 Hz and its multiples being the most prominent amplitudes envelope spectrum. The electrical current frequency (f_e) of 52.51 Hz after filtering is considerably lower, as shown in Figure 42.

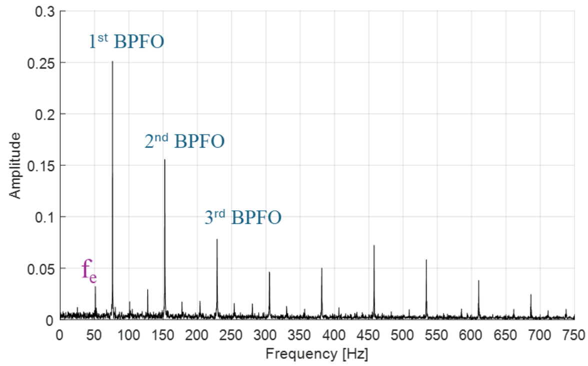


Figure 42. Envelope spectrum of sample N15_M07_F10_KA04_1 without filtering.

In this case, the leading problem is the influence of the inverter in the readings, adding noise to it, increasing its amplitude. The challenge of creating representative signals is training the model to distinguish health operation even with the possible interference of other components and contamination of the signal.

5.4 Nested cross-validation – Real data

The replication of the identification with the same samples, in a combination of training with three bearings sets and testing two, gives 10 train-test groups, as shown in Table 4. For each run, the train set was cross-validated using a Bayes Optimization, to select the best parameters for the model, by applying the feature selection of the encoder and optimizing its parameters simultaneously with the SVM.

Table 4. Data set used in the nested cross-validation.

Group	Healthy	Outer race damage	Inner race damage
0	K001	KA04	KI04
1	K002	KA15	KI14
2	K003	KA16	KI16
3	K004	KA22	KI18
4	K005	KA30	KI21

The bounds for the optimization are Δ_{fnh} between 0.35 and 0.80, the number of features N are set to be between 50 and 100, SVM hyperparameter γ is between 0.01 and 0.6, and C between 1 and 8. It uses the kernel poly of degree 6 with independent coefficient $r = 1.5$. Table 5 presents the scores and hyperparameters tunned for each fold.

Table 5. Nested cross validation result and Hyperparameter of each fold

Fold	Accuracy [%]		Groups		Encode		SVC	
	Validation	Test	Training	Test	Δfnh	N	C	γ
0	83.3	79.2	[2 3 4]	[0 1]	0.44	100	8.00	0.60
1	79.2	95.2	[1 3 4]	[0 2]	0.44	50	1.00	0.60
2	81.0	90.0	[1 2 4]	[0 3]	0.47	50	8.00	0.60
3	78.3	86.3	[1 2 3]	[0 4]	0.49	50	1.00	0.60
4	76.3	77.1	[0 3 4]	[1 2]	0.45	100	1.00	0.39
5	85.7	67.9	[0 2 4]	[1 3]	0.46	50	1.00	0.50
6	91.5	53.3	[0 2 3]	[1 4]	0.78	50	4.06	0.60
7	78.2	87.9	[0 1 4]	[2 3]	0.48	88	6.70	0.60
8	67.1	71.5	[0 1 3]	[2 4]	0.73	100	8.00	0.39
9	84.9	83.1	[0 1 2]	[3 4]	0.47	94	1.12	0.38
Mean	**80.6**	**79.2**	-	-	-	-	-	-

The influence of size and variability of the training set on the accuracy can be demonstrated by the slightly higher mean value of accuracy in the train-validation score, 80.6 %, versus 79.2% in the test score. The great difference indeed is in the variability of the test set accuracy, which characterizes the influence of the training data in the final result, as well as the data being tested.

When groups 1, 3 and 4 are the majority in the training set, the validation score is from 10 to 20% lower than the test score, folds 1, 2, 3 and 7. However, training with sets 2, 3, and 0 and testing with 1 and 4, leads to a test accuracy of 53.3%, significantly lower than the second-last accuracy, 71.5%, and the mean accuracy values for validation and test. The confusion matrix for all the 10 folds is presented in Figure 43.

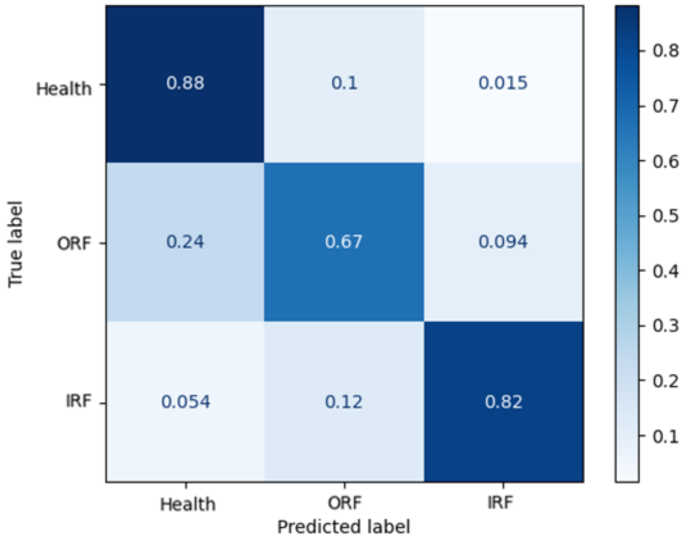


Figure 43. Confusion matrix of the nested cross-validation.

To assess the behavior of each bearing group individually, Figure 44 demonstrates each fold of the cross-validation when the model is trained with just one group of 240 samples, to identify the role of each group in the outcome. The internal cross-validations, to optimize the hyperparameters, were divided into four groups based on the operational conditions, to avoid patter memorization.

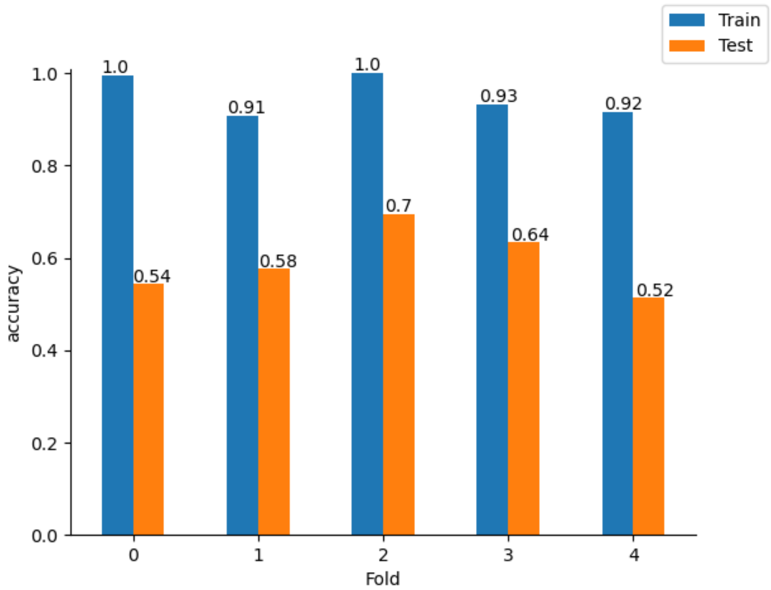


Figure 44. Training with just one group, optimizing with cross validation dividing in groups per condition.

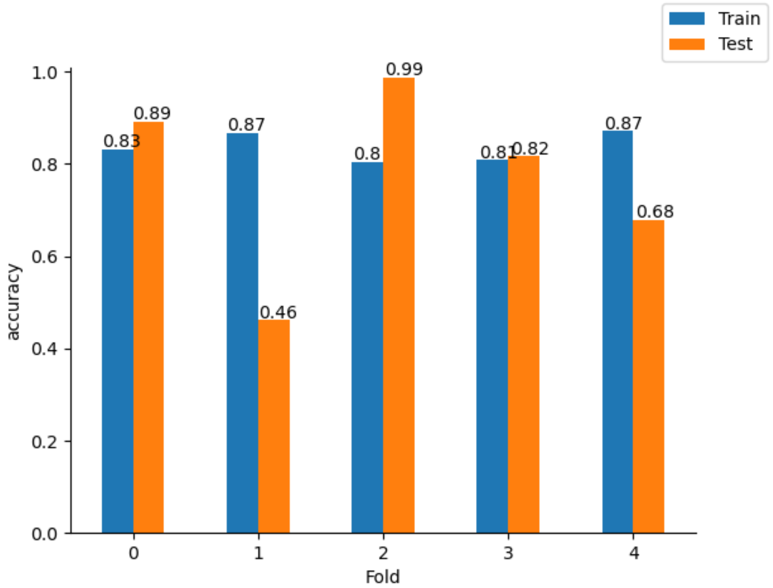


Figure 45. Training with four groups and testing with just one.

Figure 45 shows the opposite of Figure 44. The model is trained with four groups and tested with just one. The average test accuracy of the one group trained is 59.6 %, while, testing with just one group led to an average of 76.8% accuracy, near the condition in Table 5**Erro! Fonte de referência não encontrada.**, 79.15 %, but smaller. This can be due to the natural variation expected in the process, the outcome of a model depends on more variables than just the sample size.

5.5 Training set with simulated faults

The model used here was designed to train and extract features simultaneously, extending the effort to avoid under or overfitting to this usually pre-processing phase. To target these issues, the simulated training set was built in groups with non-intersection characteristics such as the extent and depth of the spall, as well as uncertainties unrelated to the faulty, such as noise nature and level, small fluctuations in the rotational speed and the unbalanced force, and clearance or interference in the in REB assembly.

A biased model, intolerant to noise, could be brought by a training set with low or meaningless variations or could be a product of an optimization that learned specificities of the models, memorizing patterns poorly correlated to the fault itself. Dividing the simulation data set into groups enables the corrected pattern identification. Otherwise, the optimization can rapidly converge to 100% accuracy in the inner loop, when the inner training and validation set have pairs of similar samples that converge to the right label, independent of the hyperparameters selection, and the final model would be underfitted.

Table 6. Dataset construction based on simulation.

group	me [kg.m]	real signal ratio	angular span [º]	max depth ratio
1	1.0E-04	0.7	1.44	0.90
				0.85
				0.80
			1.80	0.90
				0.85
				0.80
			2.16	0.90
				0.85
				0.80
2	1.0E-08	0.9	7.20	0.40
				0.30
				0.20
			12.6	0.40
				0.30
				0.20
			18.0	0.40
				0.30
				0.20

Table 6 shows the dataset generated with three samples for each simulation, randomly variating the added white noise level around the -5 dB and 0 dB of signal-to-noise ratio (SNR), It is necessary to keep the SNR low, once the fault amplitude is practically the only source of excitation in acceleration signal.

The test data set is similar to that applied in the previous chapter in Table 4. The health bearing K001 is substituted by K006 since K001 is going to be part of the training set, as presented in Table 7.

Table 7. Test data set for the model trained with simulated damages.

Group	Healthy	Outer race damage	Inner race damage
0	K006	KA04	KI04
1	K002	KA15	KI14
2	K003	KA16	KI16
3	K004	KA22	KI18
4	K005	KA30	KI21

The optimization was carried out with 20 points, and the best parameter were $\Delta_{fnh} = 0.48$, $N = 44$, with an overlap, $\sigma = -1.0$ for the encoder. The best parameters for the SVM were a polynomial kernel, degree 5 with the independent coefficients $r = 1.05$, C = 4.40 and $\gamma = 0.04$. The validation score, inner test, is 92.4 % accuracy.

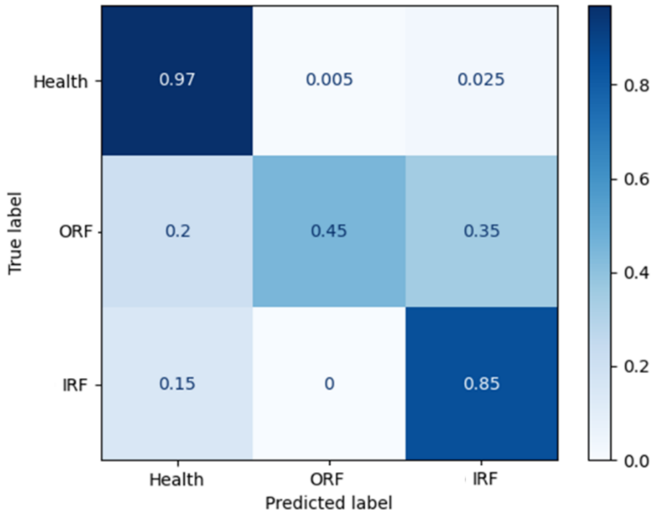


Figure 46. Confusion matrix, training with simulated and testing with real data.

Figure 46 shows the confusion matrix for a model trained with simulated fault signals only and tested with all samples, as referenced in Table 7, resulting in an accuracy of 76% for the totality of 1200 samples. An accuracy of 97% is expected for health label, once the training set uses real data for undamaged training, this can also cause the precision to be 73% and 27% of the health labeled samples were in fact damaged ones. The misleading prediction of the real data can be explained due to other characteristics that are more prominent than the fault itself.

The outer race fault has the lowest accuracy, only 45%, and the highest precision close to 100%. The model is good in representing and differentiating the OR fault but is not enough comprehensive to represent all possible outer race faults. 35% of those samples were classified as IR fault, which is acceptable but not ideal, once the location of the fault is relevant information.

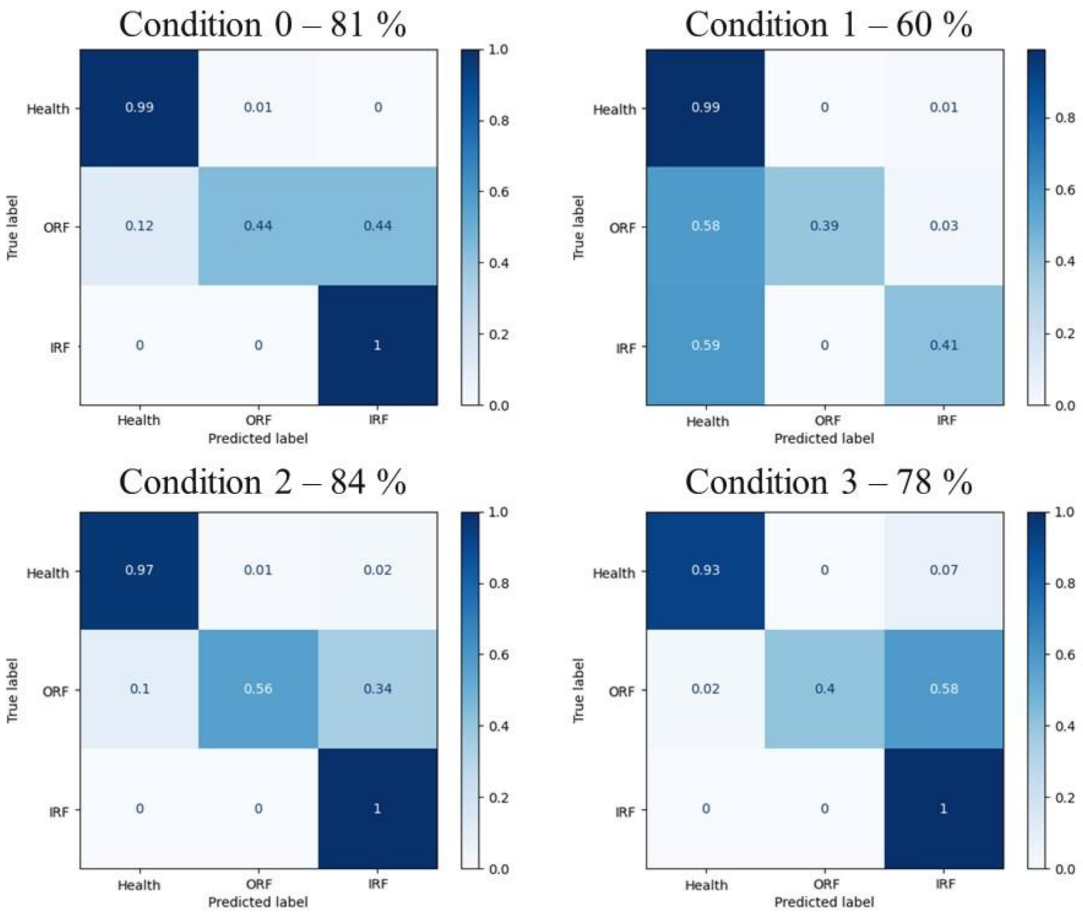


Figure 47. Prediction per condition - confusion matrices.

The predictions for the same trained model of Figure 46 are now presented in Figure 47 divided by operational conditions. While the samples with 1500 RPM, conditions 0, 2 and 3, achieve accuracy around 80%. with the misclassified cases mostly ORF and mainly mistaken as IRF. The confusion matrix for the predictions for the 900 RPM, condition 1, shows not only a poor outcome of 60% accuracy but also a 60% misclassification of damaged bearings as health.

In Figure 48, the model train and test used conditions 0 and 1 individually. In both cases, there are no significant changes in the accuracy. It can be concluded that is a valid option to train only one model with more than one condition, in similar proportions. The poor identification in the 900 RPM can be attributed to lower amplitudes, as it was observed even in health bearing, in Figure 33. It may lead to fault signals in lower velocities approaching a health case result.

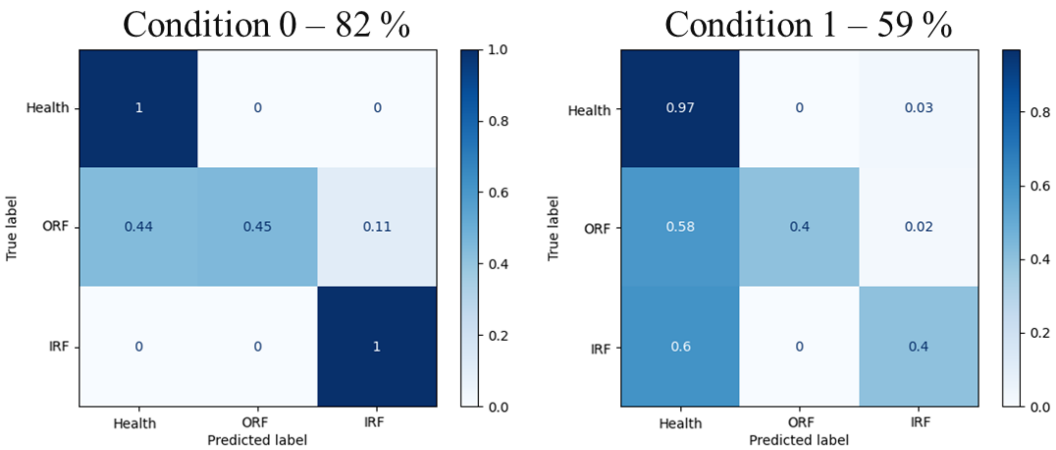


Figure 48. Training and testing the conditions separately.

Figure 49 shows the confusion matrix of a model trained and tested only with 1500 RPM samples, but with two load conditions, 1000N and 400N, respectively, conditions 0, 2 and 3 in Table 2. The total accuracy with the training data set is 82%, and again the health and the IRF samples scored practically 100% of the ORF and 47%.

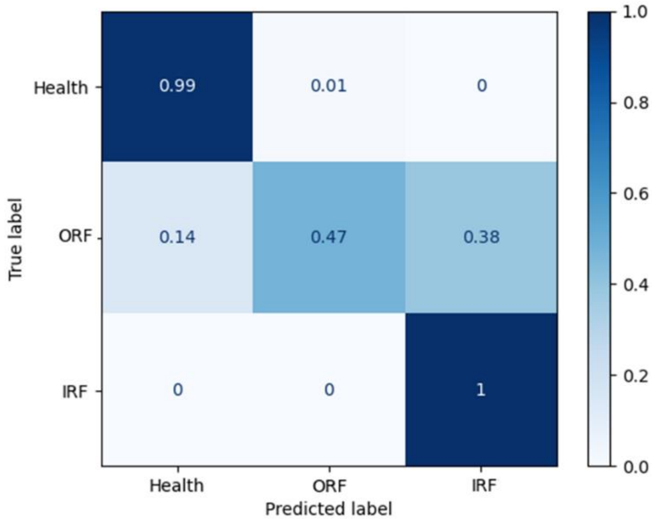


Figure 49. Confusion Matrix of identification of model 1500 RPM.

Table 8 shows the details of the identification of the confusion matrix of Figure 49, the level of the damage is described accordingly in the files of the profile of rolling bearing damage, as the length and description, are provided together with the respective data.

Table 8. Description of tested damaged rolling elements bearings.

REB	Accuracy	Level	Length	Damage description
KA04	1.00	1	2 mm	Single point pitting
KA15	0.00	1	<1 mm	Single particle-caused indentation
KA16	0.93	2	3 mm	2 pitting damages, with 2 mm and 3 mm length
KA22	0.00	1	<2 mm	Single point pitting
KA30	0.43	1	<1 mm	Distributed and random particle-caused indentations
KI04	1.00	1	2 mm	Multiple damage: pitting and particle-caused indentations.
KI14	1.00	1	1 mm	Multiple damage, pitting and particle-caused indentations.
KI16	1.00	3	6 mm	Single point pitting
KI18	1.00	2	2.5 mm	Single point pitting
KI21	1.00	1	1 mm	Single point pitting

The accuracy per condition is considerably homogeneous, namely, condition 0 with 82%, condition 2 with 85%, and 3 with 80 %. Conditions 0 and 2 are almost the same, considering the torque is the only difference.

The superiority in the inner race fault identification can be attributed to some factors: the inner race fault produces more intense impacts once the geometry of opposite curvatures leads to a deeper and faster passage over the spall. The system rotational frequency is modulated by its fault, independent of the unbalanced level. And the majority of the inner race faults are already higher level.

All IR fault bearings were 100% correctly identified, as the bearing KA04, with OR fault, and the health bearing, except K005, was 95% correctly identified. The misclassified bearings are presented in Table 9.

Table 9. Misclassified bearings descriptions and identification.

REB	Accuracy	Identified as			Level	Length	Description
		HB	OR	IR			
KA15	**0.00**	0.63	**0.00**	0.37	1	<1 mm	Single particle-caused indentation
KA16	**0.93**	0.02	**0.93**	0.05	2	3 mm	Pitting: 2 mm and 3 mm length
KA22	**0.00**	0.03	**0.00**	0.97	1	<2 mm	Single point pitting
KA30	**0.43**	0.03	**0.43**	0.53	1	<1 mm	Distributed particle-caused indentations
K005	**0.95**	**0.95**	0.05	0.00	0	-	-

From the five OR fault datasets evaluated, two were properly identified, KA04 and KA16, the last with 93% accuracy. The three faults KA15, KA22 and KA30 were poorly scored. All of those REB are level 1 damage and have distinguished fault characteristics. Among those three REB, KA30 had the best score, it had 96% of samples corrected were identified as a damaged bearing. Even though the type of the fault could not be effectively predicted, it scored close to a dummy classifier, with 50% each.

The problematic results come from KA22 and KA15 bearings, being not a single sample corrected classified. However, in distinct ways, KA15 had 63% of samples identified as healthy, the other 37% as IR fault, while KA22 had nearly all samples, 97%, misclassified as IR fault. And, although, KA22 is at the boundary of level 1 fault and described just as KA04, that last scored 100%. One consideration to be taken into account is the source misclassifying these bearings is beyond the fault size only.

Data from KA22 and KA15 bearings are challenging to identify by analyzing the signal manually, as spotted in Figure 50. The frequencies that can be identified are the electric current and its multiples, and multiples of 144 Hz, which does not correspond to none of the critical frequencies expected from the bearing 6203. The signal in Figure 50 is from KA22 bearing, sample 1 in condition 0, in which the shaft natural frequency is 24.9 Hz. The multiple frequencies of 144.5 Hz, an unknown source, correspond to the highest amplitudes in the complete frequency band. In addition, in this spectrum, the BPFO of 76.5 Hz, does not even point out, nor its multiple frequencies.

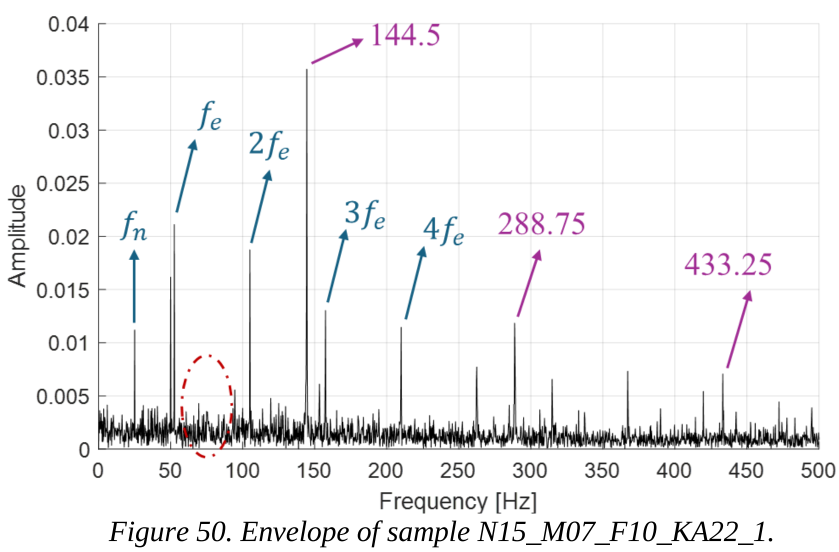


Figure 50. Envelope of sample N15_M07_F10_KA22_1.

The identification relies on frequency patterns of the same bearing type and geometry, it can deal with other frequency amplitudes, as long as some of the fault characteristic frequencies arise in the signal, and this does not occur in KA22 bearing data. The other two unsuccessful identifications are samples from bearings KA15 and KA30. In Figure 51 (a) from bearing KA15, the BPFO is present, although incipient compared to the other frequencies, and in Figure 51 (b) for bearing KA30, the BPFO is more evident, although still less than the other frequencies. In those two bearings, the type of fault is particle-indentation, with smoother fault edges, which do not excite the system as strongly as pitting/spalling fault.

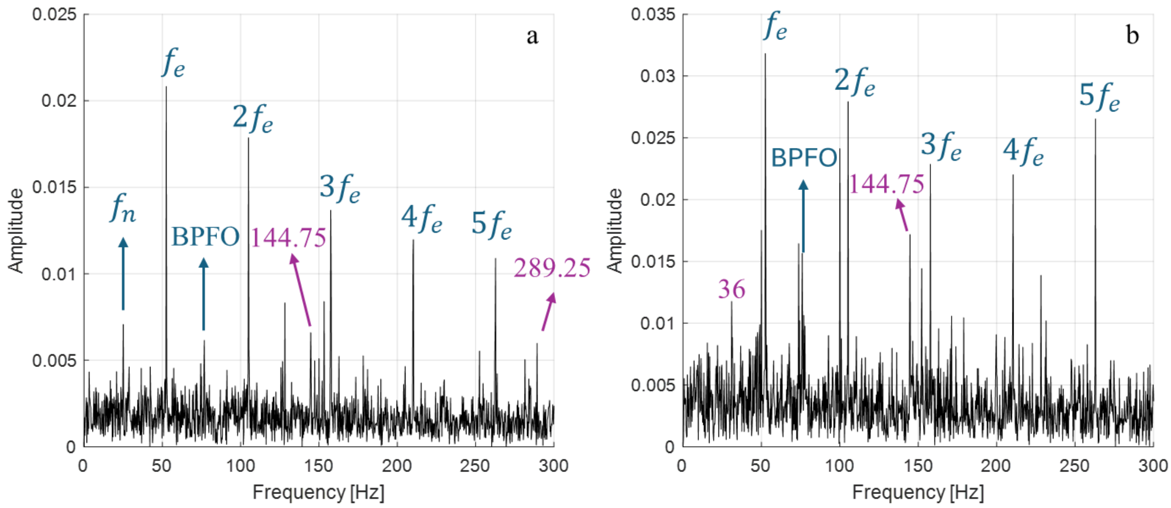


Figure 51. Envelope spectrum from (a) N15_M07_F10_KA15_1 and (b) sample N15_M07_F10_KA30_1.

While the unidentified frequency of 144 Hz and multiples can be spotted in different signals, as in Figure 51 (a) and (b), and most likely yielded by other components of the system, the multiples of this frequency are more evident than any other in KA22, Figure 50.

Another concern is the system rotational frequency, f_n, in KA22 and KA15. The rotational frequency should not be a factor that modulates the REB signal in case of an outer race fault. This behavior suggests other abnormal conditions, such as an inadvertently inner race fault, or an inadequacy in the assembly, which was not all deflected by the self-aligned ball bearing into the outer housing, as in the test bearing description by Lessmeier *et al.* (2014).

5.6 Data augmentation using simulated faults

This section accesses the improvement of the model by adding simulated signals to the data set, in the way that the model is trained by one group of real data set, one on each fold of the nested cross-validation. This model scored 80%, 79%, and 81% per conditions 0, 2 and 3, respectively, and 65% for condition 1, trained with 592 simulation samples, 296 IR and 296 OR, a total of 832, and 960 tested.

Table 10. Augmentation nested cross-validation.

Fold	**Validation Score**	**Test Score**	**Real training set added**	σ	Δ_{fnh}	**N**	**C**	γ
0	0.86	0.80	K005 KA30 KI21	-0.92	0.48	56	2.06	5E-03
1	0.85	0.87	K004 KA22 KI18	-0.95	0.49	60	2.00	5E-02
2	0.97	0.66	K003 KA16 KI16	-0.67	0.35	45	5.00	1E-03
3	0.84	0.83	K002 KA15 KI14	-0.88	0.45	53	4.98	3E-01
4	0.95	0.66	K006 KA04 KI04	-0.71	0.36	74	4.69	9E-02

In Table 10 the results are presented with the validation scores at the optimized parameters, and how the characteristics of each data set affect the choice of the encoder parameters. Folds 2 and 4 have narrow Δ_{fnh} after the optimization, for having more severe faults in the training set the main frequencies amplitude are more prominent than other peaks, reaching higher validation scores, only achieving 0.66 accuracy of the test set.

Table 11 shows the prediction for each bearing group being tested versus the model tested with simulated faults and one bearing group. The total accuracy remains 76% between training the model with only simulated faults and the training the model with the same simulated data set in addition to one group of real data.

An evident enhancement in the overall accuracy of a model trained with only one group of real faults, as presented in Figure 44, improved from 0.59 to 0.76 in the mean accuracy. The result is particularly improved when training with KA22 or KA15, this can be credited to the similarity between those bearings signals, expanding the detection for this case. This is represented in Figure 52, compering the evolution of the overall accuracy by augmentating the training data set with the simulated fault data.

Table 11. Individualized score per bearing group with mixed simulation and real fault nested validation

REB	train / test	0	1	2	3	4	Score per REB tested	Score per group tested
K006	**0**	-	0.95	0.96	1.00	0.99	**0.98**	**0.92**
KA04		-	0.99	0.99	0.99	0.99	**0.99**	
KI04		-	0.75	0.93	0.75	0.75	**0.79**	
K002	**1**	0.71	-	0.95	0.66	0.79	**0.78**	**0.62**
KA15		0.00	-	0.00	0.76	0.18	**0.23**	
KI14		0.91	-	0.93	0.75	0.76	**0.84**	
K003	**2**	0.99	1.00	-	1.00	1.00	**1.00**	**0.99**
KA16		1.00	1.00	-	0.99	0.86	**0.96**	
KI16		1.00	1.00	-	1.00	1.00	**1.00**	
K004	**3**	0.71	0.78	0.54	-	1.00	**0.76**	**0.67**
KA22		0.00	0.79	0.00	-	0.26	**0.26**	
KI18		1.00	1.00	1.00	-	1.00	**1.00**	
K005	**4**	0.71	0.36	0.76	0.91	-	**0.69**	**0.62**
KA30		0.00	0.40	0.04	0.84	-	**0.32**	
KI21		0.84	0.96	0.88	0.74	-	**0.85**	
Score REB training		**0.66**	**0.83**	**0.66**	**0.87**	**0.80**	**0.76**	**0.76**

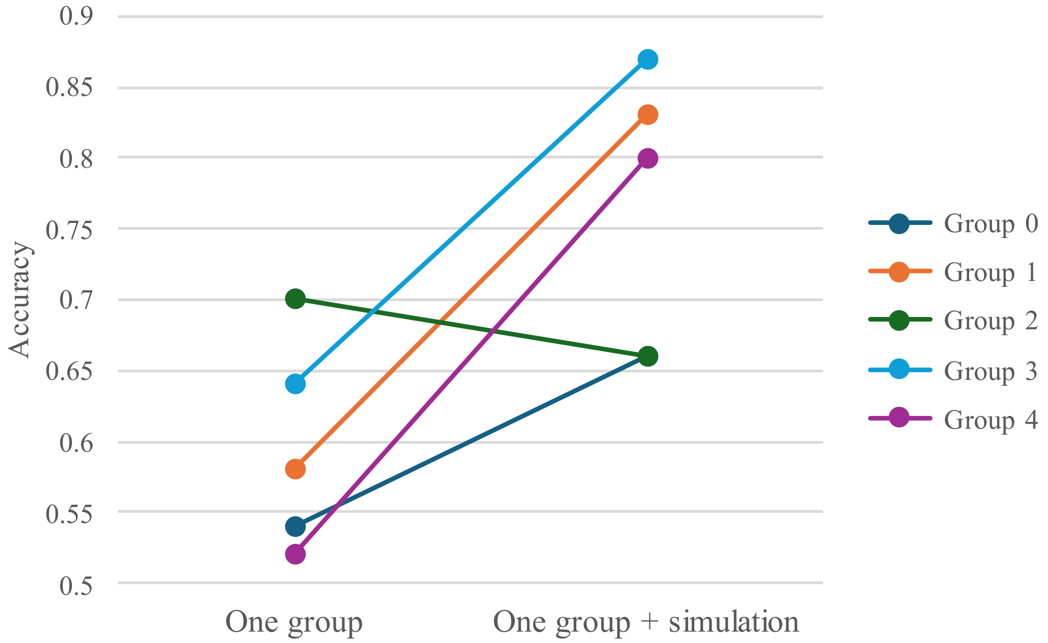


Figure 52. Comparison between the overall result in two cases: training with just one group of real fault and training with the same group adding the simulation data set.

5.7 Learning curve and convergence

The definition of the size of the training set depends on how broad it is intended to be the testing. If the application conditions are known, it is necessary to consider variations in radial clearance, temperature, and spall dimensions.

The maximum and minimum values are defined based on the extreme conditions assumed to occur, and the points and combinations in between depend on the evaluation of the learning curve. This relates to the amount of training data necessary to obtain the optimum result, given the same conditions, and up to what point the model accuracy saturates if it continuously increases in complexity as the number of samples.

Observing the learning curve for this case, clearly after 400 samples, the variability of the testing data set does not depend on the number of training samples. On the other hand, the validation score rapidly converges to 100%. For that reason, training for hyperparameters tunning was done by dividing the training set into two groups with differences in fault severity and other operational conditions as shown in Table 7.

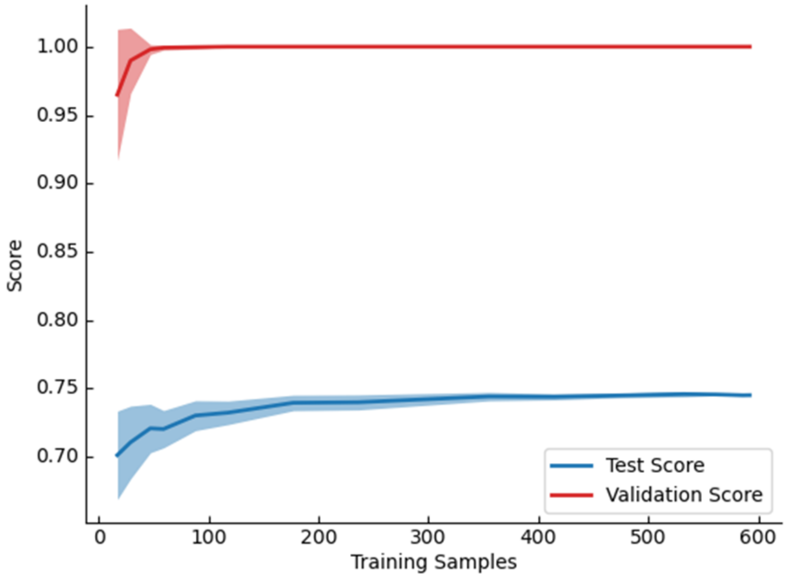


Figure 53. Learning curve.

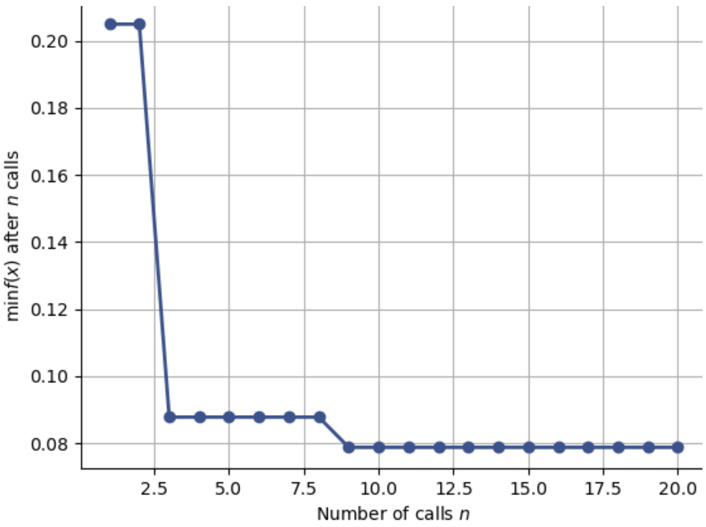


Figure 54. Error convergence per call of the optimization function on the training-validation data set.

Another variable in the optimization is the number of points it needs to explore to find the optimal solution. Taking as an example, the optimizations of the model tested in the confusion matrix of Figure 46, the Bayesian optimization rapidly converges to one minimal region up to 20 function calls, as shown in Figure 54, and the error of 0.072 is practically constant above 10 calls. The values of 20 exploitation points for simulation intended to avoid overfitting the model, which could occur if the accuracy of the training/validation set were close to 100%.

6 Conclusions

The purpose of this work is to explore the possibility of creating a training data set with fault simulation that is capable of training a machine learning model for the identification and classification of real measured rolling element bearing fault signals. The feature selection method that encodes and scales the envelope signal throughout optimization was demonstrated to be a promising tool for fault classification. This work approach explored the build of the training set data, and presented a new way set of feature selection beyond statistical moments, by introducing the idea of encoding, to be interpreted by ML algorithms.

The fault REB model showed a coherent response compared to the literature of the work of Sawalhi and Randall (2011). The same model generates the data set of the Paderborn test bench to train the SVM algorithm with satisfactory results in identifying faults with different profiles, such as debris indentation and pitting, but also multiple faults in the same bearing surface.

The results have demonstrated that is not only possible to identify the REB condition with 76% accuracy in this dataset, but also improve from 59 % to 79% the outcome when the training set is composed of just one bearing and is augmented by simulated data. This could improve health monitoring of machines that are not simple to disassemble and do not have a history of measured faults.

It is clear from the results, that the rotating speed significantly alters the amount of information on the health condition passed by the envelope signal and, therefore, the quality of the training set and the possibility of fault identification. The test set of 1500 RPM had 82% of its samples correctly identified. Of those, health and inner race fault samples, had almost 100% accuracy, and the outer race fault samples scored around 47%. The explanation may be a consequence of weaker impacts of outer race fault in combination with other abnormal operation conditions and can be even considered the presence of other undetected faults, as one of the bearings had been misclassified as inner race fault by 97%.

Another keen factor in the identification is the electrical current influence in the Paderborn data set. It is caused by the frequency inverter and defied the possibility of automatized filtration or band demodulation. For that reason, the choice of applying the Hilbert transform to the whole band spectrum produced better results. Another solution presented was adding the health signals with meanly this electrical frequency influence to bring the simulation signal closer to the measured fault, and it turned out to be a reasonable solution to build the training set.

The methodology and conclusions from this thesis contribute to future research and application of simulation data sets to build machine learning models. It presents all the steps for developing the numerical simulations of a real system, characterizing the ball bearings reduced EHD force model, introducing fault as imposed displacement and velocity smooth functions. As well as incorporate these simulated signals with a proposed encoded method for extracting and selecting features, altogether with the SVM algorithm optimization. The final solution is a machine learning model able to identify and classify the health status of damaged rolling bearings.

References

AINI, R. **Vibration monitoring and modelling of shaft/bearing assemblies under concentrated elastohydrodynamic condition**. 1990. Kingston Polytechnic, 1990.

ANTONI, J.; RANDALL, R. B. A stochastic model for simulation and diagnostics of rolling element bearings with localized faults. **Journal of Vibration and Acoustics, Transactions of the ASME**, vol. 125, no. 3, p. 282–289, 2003. https://doi.org/10.1115/1.1569940.

ANTONI, J.; RANDALL, R. B. Differential diagnosis of gear and bearing faults. **Journal of Vibration and Acoustics, Transactions of the ASME**, vol. 124, no. 2, p. 165–171, 2002. https://doi.org/10.1115/1.1456906.

ANTONI, Jérôme. Fast computation of the kurtogram for the detection of transient faults. **Mechanical Systems and Signal Processing**, vol. 21, no. 1, p. 108–124, 2007. https://doi.org/10.1016/j.ymssp.2005.12.002.

ANTONI, Jérôme. The spectral kurtosis: A useful tool for characterising non-stationary signals. **Mechanical Systems and Signal Processing**, vol. 20, no. 2, p. 282–307, 2006. https://doi.org/10.1016/j.ymssp.2004.09.001.

ANTONI, Jérôme; RANDALL, R. B. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. **Mechanical Systems and Signal Processing**, vol. 20, no. 2, p. 308–331, 2006. https://doi.org/10.1016/j.ymssp.2004.09.002.

BIZARRE, L.; NONATO, F.; CAVALCA, K. L. Formulation of five degrees of freedom ball bearing model accounting for the nonlinear stiffness and damping of elastohydrodynamic point contacts. **Mechanism and Machine Theory**, 2018. https://doi.org/10.1016/j.mechmachtheory.2018.03.001.

BIZARRE, L.; NONATO, F.; CAVALCA, K.L. Investigation on the convergence of EHD lubrication in angular contact ball bearing. 2016. **IMEchE** [...]. Manchester: IMEchE - Institution of Mechanical Engineers , 2016. p. 367–377.

BORGHESANI, P.; SMITH, W. A.; RANDALL, R. B.; ANTONI, J.; EL BADAOUI, M.; PENG, Z. Bearing signal models and their effect on bearing diagnostics. **Mechanical Systems and Signal Processing**, vol. 174, no. March, p. 109077, 2022. DOI 10.1016/j.ymssp.2022.109077. Available at: https://doi.org/10.1016/j.ymssp.2022.109077.

BROCHU, Eric; CORA, Vlad M.; DE FREITAS, Nando. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. 12 Dec. 2010. Available at: http://arxiv.org/abs/1012.2599.

CARRER, L.; BIZARRE, L.; CAVALCA, K. L. Rotor-angular contact ball bearing system study using EHD lubrication and comparison with experimental tests. **Proceedings of IMEchE VIRM 12**. Bath: IMechE, London, 2020. p. 182–191.

CERRADA, Mariela; SÁNCHEZ, René Vinicio; LI, Chuan; PACHECO, Fannia; CABRERA, Diego; VALENTE DE OLIVEIRA, José; VÁSQUEZ, Rafael E. A review on data-driven fault severity assessment in rolling bearings. **Mechanical Systems and Signal Processing**, vol. 99, p. 169–196, 2018. https://doi.org/10.1016/j.ymssp.2017.06.012.

CHRISTOPHER M. BISHOP. **Pattern Recognition and Machine Learning**. [*S. l.*: *s. n.*], 2006. https://doi.org/10.1007/978-3-030-57077-4_11.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, vol. 20, no. 3, p. 273–297, Sep. 1995. DOI 10.1007/BF00994018. Available at: http://link.springer.com/10.1007/BF00994018.

CUI, Bodi; WENG, Yang; ZHANG, Ning. A feature extraction and machine learning framework for bearing fault diagnosis. **Renewable Energy**, vol. 191, p. 987–997, 2022. DOI 10.1016/j.renene.2022.04.061. Available at: https://doi.org/10.1016/j.renene.2022.04.061.

D. DOWSON. A generalized Reynolds equation for fluid-film lubrication. **International Journal of Mechanical Science**, vol. 4, no. 2, p. 159–170, 1962. .

DARLOW, Mark S; NO, Oovt Accession; NUMBER, I Rlcipient S Catalog; DAR, Mark S; BADGLEY, Robert H; HOGG, G W. **TECHNIQUES FOR BEARING APPLICATION OF HIGH-FREQUENCY RESONANCE TECHNIQUES FOR BEARING DIAGNOSTICS IN HELICOPTER**. [*S. l.*: *s. n.*], 1974.

DOWSON, D.; HIGGINSON, G.R. **Elasto-Hydrodynamic Lubrication**. SI. Oxford: Pergamon Press, 1977. DOI 10.1016/C2013-0-05764-7. Available at: https://linkinghub.elsevier.com/retrieve/pii/C20130057647.

DWYER, Roger F. Detection of Non-Gaussian Signals By Frequency Domain Kurtosis Estimation. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**, vol. 2, p. 607–610, 1983. https://doi.org/10.1109/icassp.1983.1172264.

E. MELDAU. Die Bewegung der Achse von Wälzlagern bei geringen Drehzahlen. **Werkstatt und Betrieb**, vol. 84, no. 7, p. 308–313, 1951. .

ENDO, Hiroaki. THE UNIVERSITY OF NEW SOUTH WALES School of Mechanical and Manufacturing Engineering Simulation of Gear Faults and Its Application to the Development of Differential Diagnostic Technique. no. July, 2005. .

FREITAG, Edgar. **Encyclopedia of Lubricants and Lubrication**. [*S. l.*: *s. n.*], 2014. https://doi.org/10.1007/978-3-642-22647-2.

FUKATA, Satoru; GAD, Emil Halim; KONDOU, Takahiro; AYABE, Takashi; TAMURA, Hideyuki. On the Radial Vibration of Ball Bearings : Computer Simulation. **Bulletin of JSME**, vol. 28, no. 239, p. 899–904, 1985. DOI 10.1299/jsme1958.28.899. Available at: http://www.jstage.jst.go.jp/article/jsme1958/28/239/28_239_899/_article.

GARDNER, Jacob R.; KUSNER, Matt J.; XU, Zhixiang; WEINBERGER, Kilian Q.; CUNNINGHAM, John P. Bayesian optimization with inequality constraints. **31st International Conference on Machine Learning, ICML 2014**, vol. 3, p. 2581–2591, 2014. .

GUO, Lei; CHEN, Jin; LI, Xinglin. Rolling bearing fault classification based on envelope spectrum and support vector machine. **JVC/Journal of Vibration and Control**, vol. 15, no. 9, p. 1349–1363, 2009. https://doi.org/10.1177/1077546308095224.

GUPTA, P. K. Dynamics of Rolling-Element Bearings—Part I: Cylindrical Roller Bearing Analysis. **Journal of Lubrication Technology**, vol. 101, no. 3, p. 293–302, 1 Jul. 1979a. https://doi.org/10.1115/1.3453357.

GUPTA, P. K. Dynamics of Rolling-Element Bearings—Part II: Cylindrical Roller Bearing Results. **Journal of Lubrication Technology**, vol. 101, no. 3, p. 305–311, 1 Jul. 1979b. https://doi.org/10.1115/1.3453360.

GUPTA, P. K. Dynamics of Rolling-Element Bearings—Part III: Ball Bearing Analysis. **Journal of Lubrication Technology**, vol. 101, no. 3, p. 312–318, 1 Jul. 1979c. https://doi.org/10.1115/1.3453363.

GUPTA, P. K. Dynamics of Rolling-Element Bearings—Part IV: Ball Bearing Results. **Journal of Lubrication Technology**, vol. 101, no. 3, p. 319–326, 1 Jul. 1979d. https://doi.org/10.1115/1.3453364.

GUPTA, P. K. Transient Ball Motion and Skid in Ball Bearings. **Journal of Lubrication Technology**, vol. 97, no. 2, p. 261–269, 1 Apr. 1975. https://doi.org/10.1115/1.3452568.

GUPTA, P. K.; DILL, J. F.; BANDOW, H. E. Dynamics of Rolling Element—Bearings Experimental Validation of the DREB and RAPIDREB Computer Programs. **Journal of Tribology**, vol. 107, no. 1, p. 132–137, 1 Jan. 1985. https://doi.org/10.1115/1.3260989.

GUPTA, Pankaj; PRADHAN, M. K. Fault detection analysis in rolling element bearing: A review. **Materials Today: Proceedings**, vol. 4, no. 2, p. 2085–2094, 2017. DOI 10.1016/j.matpr.2017.02.054. Available at: http://dx.doi.org/10.1016/j.matpr.2017.02.054.

GUPTA, Pradeep K. **Advanced Dynamics of Rolling Elements**. New York, NY: Springer New York, 1984. https://doi.org/10.1007/978-1-4612-5276-4.

H. PERRET. Elastische Spielschwingungen konstant belasteter Wälzlager. **Werkstatt und Betrieb**, vol. 83, no. 8, p. 354–358, 1950. .

HAKIM, Mohammed; OMRAN, Abdoulhdi A.Borhana; AHMED, Ali Najah; AL-WAILY, Muhannad; ABDELLATIF, Abdallah. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. **Ain Shams Engineering Journal**, vol. 14, no. 4, p. 101945, 2023. https://doi.org/10.1016/j.asej.2022.101945.

HARRIS, T. A. **Rolling Bearing Analysis**. New York: John Wiley & Sons, 1991.

HARRIS, T A; MINDEL, M H. **ROLLING ELEMENT BEARING DYNAMICS**. [*S. l.*]: Elsevier Sequoia S.A, 1973.

HEAD, T.; KUMAR, M.; NAHRSTAEDT, H.; LOUPPE, G.; SHCHERBATYI, I. scikit-optimize/scikit-optimize. Jul. 2020. .

HERTZ, H. On the contact of elastic solids. **Hertz's Miscellaneous Papers**, , p. 146–162, 1896. .

HOWARD, Ian. **A Review of Rolling Element Bearing Vibration "Detection, Diagnosis and Prognosis."** [*S. l.*: *s. n.*], 1994.

IGARASHI, Teruo; HAMADA, Hiroyoshi. Studies on the Vibration and Sound of Defective Rolling Bearings : First Report : Vibration of Ball Bearings with One Defect. **Bulletin of JSME**, vol. 25, no. 204, p. 994–1001, 1982. https://doi.org/10.1299/jsme1958.25.994.

JOHNSON, K. L. **Contact mechanics**. 1st ed. [*S. l.*]: Cambridge University Prass, [*s. d.*].

JONES, A. B. A General Theory for Elastically Constrained Ball and Radial Roller Bearings Under Arbitrary Load and Speed Conditions. **Journal of Basic Engineering**, vol. 82, no. 2, p. 309–320, 1 Jun. 1960. DOI 10.1115/1.3662587. Available at: https://asmedigitalcollection.asme.org/fluidsengineering/article/82/2/309/442491/A-General-Theory-for-Elastically-Constrained-Ball.

KIMOTHO, James Kuria; SEXTRO, Walter. An approach for feature extraction and selection from non-trending data for machinery prognosis. **Proceedings of the Second European Conference of the Prognostics and Health Management Society**, vol. 5, no. 4, p. 1–8, 2014. .

KRAUS, J.; BLECH, J. J.; BRAUN, S. G. In Situ Determination of Rolling Bearing Stiffness and Damping by Modal Analysis. **Journal of Vibration and Acoustics**, vol. 109, no. 3, p. 235–240, 1 Jul. 1987. https://doi.org/10.1115/1.3269426.

LEI, Yaguo; LIN, Jing; HE, Zhengjia; ZI, Yanyang. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. **Mechanical Systems and Signal Processing**, vol. 25, no. 5, p. 1738–1749, 2011. DOI 10.1016/j.ymssp.2010.12.011. Available at: http://dx.doi.org/10.1016/j.ymssp.2010.12.011.

LEI, Yaguo; LIN, Jing; HE, Zhengjia; ZUO, Ming J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. **Mechanical Systems and Signal Processing**, vol. 35, no. 1–2, p. 108–126, Feb. 2013. DOI 10.1016/j.ymssp.2012.09.015. Available at: http://link.springer.com/10.1007/978-3-319-56126-4_10.

LEI, Yaguo; YANG, Bin; JIANG, Xinwei; JIA, Feng; LI, Naipeng; NANDI, Asoke K. Applications of machine learning to machine fault diagnosis: A review and roadmap. **Mechanical Systems and Signal Processing**, vol. 138, p. 106587, 2020. DOI

10.1016/j.ymssp.2019.106587. Available at: https://doi.org/10.1016/j.ymssp.2019.106587.

LESSMEIER, Christian; ENGE-ROSENBLATT, Olaf; BAYER, Christian; ZIMMER, Detmar. Data Acquisition and Signal Analysis from Measured Motor Currents for Defect Detection in Electromechanical Drive Systems. **European Conference of the Prognostics and Health Management Society 2014**, no. July, p. 1–10, 2014. DOI 10.13140/2.1.3499.3289. Available at: http://www.phmsociety.org/events/conference/europhm/14/proceedings.

LESSMEIER, Christian; KIMOTHO, James Kuria; ZIMMER, Detmar; SEXTRO, Walter. 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. **Third European Conference of the Prognostics and Health Management Society 2016**, no. Cm, p. 152–156, 2016. .

LIM, T.C.; SINGH, R. Vibration transmission through rolling element bearings, part I: Bearing stiffness formulation. **Journal of Sound and Vibration**, vol. 139, no. 2, p. 179–199, Jun. 1990a. https://doi.org/10.1016/0022-460X(90)90882-Z.

LIM, T.C.; SINGH, R. Vibration transmission through rolling element bearings, part II: System studies. **Journal of Sound and Vibration**, vol. 139, no. 2, p. 201–225, Jun. 1990b. https://doi.org/10.1016/0022-460X(90)90883-2.

LIM, T.C.; SINGH, R. Vibration transmission through rolling element bearings. Part III: Geared rotor system studies. **Journal of Sound and Vibration**, vol. 151, no. 1, p. 31–54, Nov. 1991. https://doi.org/10.1016/0022-460X(91)90650-9.

LIM, T.C.; SINGH, R. Vibration transmission through rolling element bearings, part IV: statistical energy analysis. **Journal of Sound and Vibration**, vol. 153, no. 1, p. 37–50, Feb. 1992. https://doi.org/10.1016/0022-460X(92)90625-8.

LIN, Jing; QU, Liangsheng. Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis. **Journal of Sound and Vibration**, vol. 234, no. 1, p. 135–148, 2000. https://doi.org/10.1006/jsvi.2000.2864.

LIU, Ruonan; YANG, Boyuan; ZIO, Enrico; CHEN, Xuefeng. Artificial intelligence for fault diagnosis of rotating machinery: A review. **Mechanical Systems and Signal Processing**, vol. 108, p. 33–47, 2018. DOI 10.1016/j.ymssp.2018.02.016. Available at: https://doi.org/10.1016/j.ymssp.2018.02.016.

LUBRECHT, A. A.; TEN NAPEL, W. E.; BOSMA, R. Multigrid, An Alternative Method for Calculating Film Thickness and Pressure Profiles in Elastohydrodynamically Lubricated Line Contacts. **Journal of Tribology**, vol. 108, no. 4, p. 551–556, 1 Oct. 1986. https://doi.org/10.1115/1.3261260.

LUNDBERG, Gustaf; PALMGREN, Arvid. Dynamic Capacity of Rolling Bearings. **Journal of Applied Mechanics**, vol. 16, no. 2, p. 165–172, 1 Jun. 1949. https://doi.org/10.1115/1.4009930.

MCFADDEN, P. D.; SMITH, J. D. Vibration monitoring of rolling element bearings by the high-frequency resonance technique - a review. **Tribology International**, vol. 17, no. 1, p. 3–10, 1984. https://doi.org/10.1016/0301-679X(84)90076-8.

MCFADDEN, P. D.; TOOZHY, M. M. Application of synchronous averaging to vibration monitoring of rolling element bearings. **Mechanical Systems and Signal Processing**, vol. 14, no. 6, p. 891–906, 2000. https://doi.org/10.1006/mssp.2000.1290.

MISHRA, C.; SAMANTARAY, A. K.; CHAKRABORTY, G. Ball bearing defect models: A study of simulated and experimental fault signatures. **Journal of Sound and Vibration**, vol. 400, p. 86–112, 2017. https://doi.org/10.1016/j.jsv.2017.04.010.

NONATO, Fábio; CAVALCA, Katia L. An approach for including the stiffness and damping of elastohydrodynamic point contacts in deep groove ball bearing equilibrium models. **Journal of Sound and Vibration**, 2014. https://doi.org/10.1016/j.jsv.2014.08.011.

NONATO, Fbio; CAVALCA, Katia L. On the non-linear dynamic behavior of elastohydrodynamic lubricated point contact. **Journal of Sound and Vibration**, 2010. https://doi.org/10.1016/j.jsv.2010.05.014.

OPPENHEIM, Alan V; SCHAFER, Ronald W. **Discrete-Time Signal Processing**. 3rd ed. [*S. l.*: *s. n.*], 2010.

QIU, Hai; LEE, Jay; LIN, Jing; YU, Gang. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. **Journal of Sound and Vibration**, vol. 289, no. 4–5, p. 1066–1090, 2006. https://doi.org/10.1016/j.jsv.2005.03.007.

RANDALL, Robert B; ANTONI, Jérôme. Why EMD and similar decompositions are of little benefit for bearing diagnostics. **Mechanical Systems and Signal Processing**, vol. 192, no. March, 2023. https://doi.org/10.1016/j.ymssp.2023.110207.

RANDALL, Robert B.; ANTONI, Jrme. Rolling element bearing diagnostics-A tutorial. **Mechanical Systems and Signal Processing**, vol. 25, no. 2, p. 485–520, 2011. https://doi.org/10.1016/j.ymssp.2010.07.017.

RAUBER, Thomas Walter; DA SILVA LOCA, Antonio Luiz; BOLDT, Francisco de Assis; RODRIGUES, Alexandre Loureiros; VAREJÃO, Flávio Miguel. An experimental methodology to evaluate machine learning methods for fault diagnosis based on vibration signals. **Expert Systems with Applications**, vol. 167, 1 Apr. 2021. https://doi.org/10.1016/j.eswa.2020.114022.

SAWALHI, N.; RANDALL, R. B. Simulating gear and bearing interactions in the presence of faults. Part II: Simulation of the vibrations produced by extended bearing faults. **Mechanical Systems and Signal Processing**, vol. 22, no. 8, p. 1952–1966, 2008a. https://doi.org/10.1016/j.ymssp.2007.12.002.

SAWALHI, N.; RANDALL, R.B. Simulating gear and bearing interactions in the presence of faults. **Mechanical Systems and Signal Processing**, vol. 22, no. 8, p. 1924–1951, Nov. 2008b. DOI 10.1016/j.ymssp.2007.12.001. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0888327007002725.

SAWALHI, Nader. **Diagnostics, Prognostics and Fault Simulation**. 2007. 2007. Available at: https://www.unsworks.unsw.edu.au/permalink/f/a5fmj0/unsworks_1509.

SINGH, Jaskaran; DARPE, A. K.; SINGH, S. P. 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**, vol. 100, p. 662–693, 2018. https://doi.org/10.1016/j.ymssp.2017.06.040.

SKF. SKF. 2024. Available at: **https://www.skf.com/group/products/rolling-bearings/ball-bearings/deep-groove-ball-bearings**. .

SOORI, Mohsen; AREZOO, Behrooz; DASTRES, Roza. Internet of things for smart factories in industry 4.0, a review. **Internet of Things and Cyber-Physical Systems**, vol. 3, no. April, p. 192–204, 2023. DOI 10.1016/j.iotcps.2023.04.006. Available at: https://doi.org/10.1016/j.iotcps.2023.04.006.

SUNNERSJÖ, C. S. Varying compliance vibrations of rolling bearings. **Journal of Sound and Vibration**, vol. 58, no. 3, p. 363–373, 1978. https://doi.org/10.1016/S0022-460X(78)80044-3.

TAMURA, Akiyoshi; TANIGUCHI, Osamu. Ball Bearing Vibrations : 2nd Report, On the Subharmonic Vibration of Order 1/2 Caused by the Passing Balls. **Transactions of the Japan Society of Mechanical Engineers**, vol. 26, no. 167, p. 877–883, 1960. DOI 10.1299/kikai1938.26.877. Available at: http://www.jstage.jst.go.jp/article/kikai1938/26/167/26_167_877/_article/-char/ja/.

TAMURA, Akiyoshi; TANIGUCHI, Osamu. On the Subharmonic Vibration of the Order One-half Caused by Passing Balls in a Ball Bearing. **Bulletin of JSME**, vol. 4, no. 14, p. 193–200, 1961. DOI 10.1299/jsme1958.4.193. Available at: http://www.jstage.jst.go.jp/article/jsme1958/4/14/4_14_193/_article.

TANDON, N.; CHOUDHURY, A. An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect. **Journal of Sound and Vibration**, vol. 205, no. 3, p. 275–292, 1997. https://doi.org/10.1006/jsvi.1997.1031.

THOMSON, Brian S. **SYMMETRIC PROPERTIES OF REAL FUNCTIONS**. [*S. l.*: *s. n.*], 1994.

VAPNIK, Vladimir N. Constructing Learning Algorithms. **The Nature of Statistical Learning Theory**. New York, NY: Springer New York, 1995. p. 119–166. DOI 10.1007/978-1-4757-2440-0_6. Available at: http://link.springer.com/10.1007/978-1-4757-2440-0_6.

VENNER, C. **Multilevel solution of the EHL line and point contact problems.** 1991. 1991.

VENNER, C. H.; LUBRECHT, A. A. **Multilevel Methods in Lubrication**. vol 37. [*S. l.*]: Tribology Series. Elsevier, 2000.

VISCHI CARVALHO, Rafael. **Análise dinâmica de rolamentos de esfera**. 2010. 91 f. Universidade Estadual de Campinas, Campinas, SP, 2010. DOI

10.47749/T/UNICAMP.2010.775520. Available at: https://repositorio.unicamp.br/Busca/Download?codigoArquivo=472019.

WANG, Hui; ZHENG, Junkang; XIANG, Jiawei. Online bearing fault diagnosis using numerical simulation models and machine learning classifications. **Reliability Engineering and System Safety**, vol. 234, 1 Jun. 2023. https://doi.org/10.1016/j.ress.2023.109142.

WARD BROWN, James; CHURCHILL, Ruel V. **COMPLEX VARIABLES AND APPLICATIONS**. 9th ed. [*S. l.*]: McGraw-Hill Education, 1996.

WIJNANT, Y H. **Contact Dynamics in the field of Elastohydrodynamic Lubrication**. [*S. l.*: *s. n.*], 1998.

WIJNANT, Y.H.; WENSING, J.A.; NIJEN, G.C. THE INFLUENCE OF LUBRICATION ON THE DYNAMIC BEHAVIOUR OF BALL BEARINGS. **Journal of Sound and Vibration**, vol. 222, no. 4, p. 579–596, May 1999. DOI 10.1006/jsvi.1998.2068. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0022460X98920685.

WORDEN, Keith; STASZEWSKI, Wieslaw J.; HENSMAN, James J. Natural computing for mechanical systems research: A tutorial overview. **Mechanical Systems and Signal Processing**, vol. 25, no. 1, p. 4–111, 2011. https://doi.org/10.1016/j.ymssp.2010.07.013.

YAN, Xiaoan; JIA, Minping. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. **Neurocomputing**, vol. 313, p. 47–64, 2018. https://doi.org/10.1016/j.neucom.2018.05.002.

ZARPELON RADAELLI, Gustavo. **Determinação do carregamento dinâmico em mancais de esfera de contato angular**. 2013. 86 f. Universidade Estadual de Campinas, Campinas, SP, 2013. DOI 10.47749/T/UNICAMP.2013.905856. Available at: https://repositorio.unicamp.br/Busca/Download?codigoArquivo=473585.

ZHANG, Hengcheng; BORGHESANI, Pietro; RANDALL, Robert B.; PENG, Zhongxiao. A benchmark of measurement approaches to track the natural evolution of spall severity in rolling element bearings. **Mechanical Systems and Signal Processing**, vol. 166, no. September 2021, p. 108466, 2022. DOI 10.1016/j.ymssp.2021.108466. Available at: https://doi.org/10.1016/j.ymssp.2021.108466.

APPENDIX A

The equations adapted from Sawalhi's thesis and Sawalhi and Randall (2008) represent the system with gear and four double-row ball bearings. The system has torsional vibration and translational vibration only inside the gearbox,

being the extremities isolated by flexible coupling.

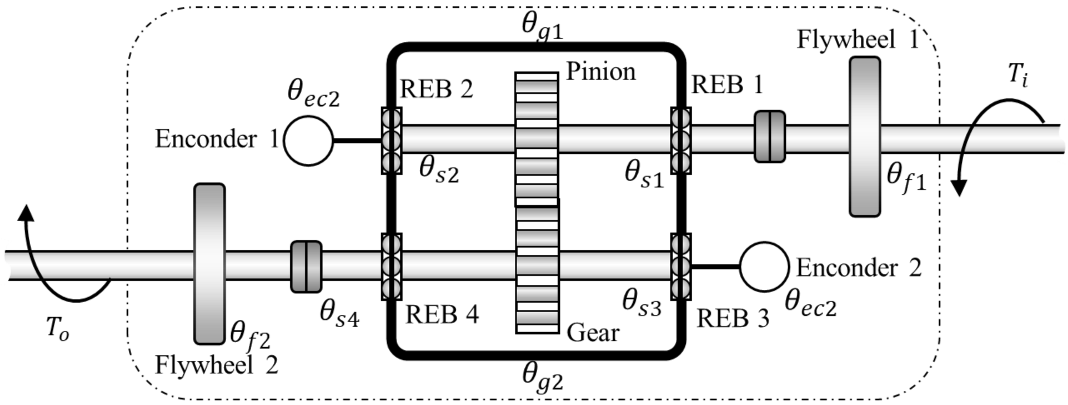


Figure 55 Scheme of the Sawalhi and Randall (2008) test rig with the angular coordinates.

$$m_{pl}\ddot{x}_{pl} + c_{pl}\dot{x}_{pl} + k_{pl}x_{pl} = f_{bxl} \tag{A1.1}$$

$$m_{pl}\ddot{y}_{pl} + c_{pl}\dot{y}_{pl} + c_{rl}(\dot{y}_{pl} - \dot{y}_{bl}) + k_{pl}y_{pl} + k_{rl}(y_{pl} - y_{bl}) = f_{byl} \tag{A1.2}$$

$$m_{sl}\ddot{x}_{sl} + k_{sl}(x_{sl} - x_{gpl}) = -f_{bxl} \tag{A1.3}$$

$$m_{sl}\ddot{y}_{sl} + k_{sl}(y_{sl} - y_{gpl}) = -f_{byl} \tag{A1.4}$$

$$m_{rl}\ddot{y}_{bl} - k_{rl}(y_{pl} - y_{bl}) - c_{rl}(\dot{y}_{pl} - \dot{y}_{bl}) = 0 \tag{A1.5}$$

Equations A1.1 to A1.5 represent the system of shaft-bearing-pedestal-mass. The subscript l goes from 1 to 4, represent the four bearings and bearings assembly. The functions f_{bx} and f_{by} are the equations for the bearing reaction forces.

Rotational degrees of freedom:

The flywheel:

$$J_{f1}\ddot{\theta}_{f1} + c_{f1}(\dot{\theta}_{f1} - \dot{\theta}_{s1}) + k_{f1}(\theta_{f1} - \theta_{s1}) = T_i \qquad \text{(A1.6)}$$

Shaft 1 (Bearing1 – pinion):

$$J_{s1}\ddot{\theta}_{s1} - c_{f1}(\dot{\theta}_{f1} - \dot{\theta}_{s1}) - k_{f1}(\theta_{f1} - \theta_{s1}) + k_{ts1}(\theta_{s1} - \theta_{gp1}) + c_{f1}(\dot{\theta}_{s1} - \dot{\theta}_{gp1}) = 0 \qquad \text{(A1.7)}$$

Pinion theta - 3

$$J_{gp1}\ddot{\theta}_{gp1} - c_{f1}(\dot{\theta}_{s1} - \dot{\theta}_{gp1}) + k_{ts2}(\theta_{gp1} - \theta_{s2}) - k_{ts1}(\theta_{s1} - \theta_{gp1}) + k_{mm}(t)r_{gp1}\delta_e + c_{mm}r_{gp1}\dot{\delta}_e = 0 \qquad \text{(A1.8)}$$

Shaft 2 (Bearing 2 – encoder 1):

$$J_{s2}\ddot{\theta}_{s2} + c_{ec1}(\dot{\theta}_{s2} - \dot{\theta}_{ec1}) - k_{ts2}(\theta_{gp1} - \theta_{s2}) + k_{ec1}(\theta_{s2} - \theta_{ec1}) = 0 \qquad \text{(A1.9)}$$

Shaft 2 (Bearing 2 - pinion):

$$J_{gp2}\ddot{\theta}_{gp2} + k_{ts3}(\theta_{gp2} - \theta_{s3}) - k_{ts4}(\theta_{gp2} - \theta_{s4}) + k_{mm}(t)r_{gp2}\delta_e + c_{mm}r_{gp2}\dot{\delta}_e = 0 \qquad \text{(A1.10)}$$

Encoder 1:

$$J_{ec1}\ddot{\theta}_{ec1} - c_{ec1}(\dot{\theta}_{s2} - \dot{\theta}_{ec1}) - k_{ec1}(\theta_{s2} - \theta_{ec1}) = 0 \qquad \text{(A1.11)}$$

Shaft 3 (Bearing 3 - gear):

$$J_{s2}\ddot{\theta}_{s3} + c_{ec2}(\dot{\theta}_{s3} - \dot{\theta}_{ec2}) + k_{ec2}(\theta_{s3} - \theta_{ec2}) - k_{ts3}(\theta_{gp2} - \theta_{s3}) = 0 \qquad \text{(A1.12)}$$

Encoder 2:

$$J_{ec2}\ddot{\theta}_{ec2} - c_{ec2}(\dot{\theta}_{s3} - \dot{\theta}_{ec2}) - k_{ec2}(\theta_{s3} - \theta_{ec2}) = 0 \qquad \text{(A1.13)}$$

Shaft 4 (Bearing 4 - gear):

$$J_{s4}\ddot{\theta}_{s4} + c_{f2}(\dot{\theta}_{s4} - \dot{\theta}_{f2}) + k_{f2}(\theta_{s4} - \theta_{f2}) + k_{ts4}(\theta_{gp2} - \theta_{s4}) = 0 \qquad \text{(A1.6)}$$

Flywheel 2:

$$J_{f2}\ddot{\theta}_{f2} - c_{f2}(\dot{\theta}_{s4} - \dot{\theta}_{f2}) - k_{f2}(\theta_{s4} - \theta_{f2}) = T_o \qquad \text{(A1.14)}$$

Translational degrees of freedom:

Pinion x:

$$m_{gp1}\ddot{x}_{gp1} + k_{s1}(x_{gp1} - x_{s1}) + k_{s2}(x_{gp1} - x_{s2}) = 0 \tag{A1.15}$$

Pinion y:

$$m_{gp1}\ddot{y}_{gp1} + k_{s1}(y_{gp1} - y_{s1}) + k_{s2}(y_{gp1} - y_{s2}) + k_{mm}(t)\delta_e + c_{mm}\dot{\delta}_e = 0 \tag{A1.16}$$

Gear x:

$$m_{gp2}\ddot{x}_{gp2} + k_{s3}(x_{gp2} - x_{s3}) + k_{s3}(x_{gp2} - x_{s3}) = 0 \tag{A1.17}$$

Pinion y:

$$m_{gp2}\ddot{y}_{gp2} + k_{s3}(y_{gp2} - y_{s3}) + k_{s4}(y_{gp2} - y_{s4}) + k_{mm}(t)\delta_e + c_{mm}\dot{\delta}_e = 0 \tag{A1.18}$$

The gear-pinion relative displacement, δ_e and velocity, $\dot{\delta}_e$, result from each shaft independent movement and the error, e_t. The cyclic stiffness throughout the contact line, $k_{mm}(t)$, is withdraw from Endo (2005) for a torque of 50 Nm and interpolated for the teeth length.

$$\begin{gathered} \delta_e = r_{gp1}\theta_{gp1} - r_{gp2}\theta_{gp2} - y_{gp1} + y_{gp2} - e_t \\ \dot{\delta}_e = r_{gp1}\dot{\theta}_{gp1} - r_{gp2}\dot{\theta}_{gp2} - \dot{y}_{gp1} + \dot{y}_{gp2} \end{gathered} \tag{A1.19}$$

The error, e_t, is present due to run-out and the toothprofile imperfections and uncertainties and is given by (Endo, 2005):

$$\begin{gathered} e_t = e_{runout} + e_{tooth} + 0.2\mathrm{randn}(t) \\ e_{runout} = 3 \times 10^{-5} \sin \Omega t \\ e_{tooth} = 1.2 \times 10^{-5} \sin \mathrm{N}\Omega t \end{gathered} \tag{A1.20}$$

The gear number of teeth is N and rand(t) gives a random distributing number at each instant.

The model also includes a slippage in the REB cage. To keep tracking of this new variant and be able to reproduce the exact same positions, it was added a DOF, related to the tested bearing cage angular position, φ_c, obtained from the numeric integration.

$$\frac{d\varphi_c}{dt} = \omega_c + \omega_{slip} \tag{A1.21}$$

and the slip velocity, ω_{slip}, between 1% and 2% of the cage velocity, is randomly generated during the integration.