

**Fault Detection and Prognostics Framework for Technical Systems Subjected
to Time-Varying Operating Conditions**

zur Erlangung des akademischen Grades
DOKTOR DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)
der Fakultät für Maschinenbau
der Universität Paderborn

genehmigte
DISSERTATION

von
Osarenren Kennedy Aimiyeagbon, M.Sc.
aus Benin City, Nigeria

Tag des Kolloquiums:	10. Juni 2026
Referent:	Prof. Dr.-Ing. habil. Walter Sextro
Korreferent:	Prof. Dr.-Ing. Gunter Kullmer

Geleitwort

Neue technische Entwicklungen entstehen heute immer öfter an den Grenzen zwischen den Disziplinen, wo verschiedene Denkweisen sich begegnen und miteinander wechselwirken. Die Entwicklung neuer mechatronischer oder intelligenter technischer Systeme setzt ein interdisziplinäres Denken und Handeln voraus.

Mechatronische Systeme bestehen aus einer mechanischen Grundstruktur, die durch Integration von Sensoren und Aktoren sowie durch informationsverarbeitende Komponenten ein hohes Maß an Flexibilität und Leistungsfähigkeit gewinnt. Die Mechatronik stellt somit eine interdisziplinäre Kombination der Fächer Maschinenbau, Elektrotechnik und Informatik dar.

Der Lehrstuhl für Dynamik und Mechatronik gehört der Fakultät für Maschinenbau der Universität Paderborn an. In Forschung und Lehre befassen wir uns mit der Modellierung, Simulation, Zuverlässigkeit, Optimierung, Betrieb, Überwachung, Diagnose und Prognose von mechanischen, mechatronischen und intelligenten technischen Systemen. Die Forschungsschwerpunkte liegen in den Bereichen Mehrkörpersysteme, Reibung, Verschleiß, Condition Monitoring, Predictive Maintenance, KI-unterstütztes Engineering, Ultraschallprozesse und -systeme. Sie spiegeln sich in den Bänden dieser Schriftenreihe, in denen Ergebnisse unserer Forschung veröffentlicht werden. Die Schriftenreihe soll dazu beitragen, den Wissenstransfer zwischen der Universität und der praktischen Anwendung zu verbessern.

Um eine zuverlässige Fehlererkennung und eine genaue Vorhersage der Restlebensdauer technischer Systeme unter zeitlich variierenden Betriebsbedingungen zu ermöglichen, stellt Herr Aimiyeagbon in dieser Arbeit ein Framework mit fünf verschiedenen Lösungsansätzen vor. Als Anwendungsbeispiele werden von Herrn Aimiyeagbon experimentelle Daten herangezogen, die jeweils zeitlich variierenden Betriebsbedingungen ausgesetzt sind. Die Ergebnisse zeigen, dass die Berücksichtigung der Betriebsbedingungen, entweder durch Merkmale, die unabhängig von den Betriebsbedingungen sind, oder durch die Einbindung der Betriebsbedingungen als zusätzliche Eingangsdaten in maschinelle Lernverfahren, die Vorhersagegüte von Fehlererkennungs- und Prognosemethoden verbessert.

Paderborn, den 18. Juni 2026

Prof. Dr.-Ing. habil. Walter Sextro

Acknowledgment

The completion of this thesis would not have been possible without the support, encouragement, and contributions of many people, to whom I would like to express my sincere gratitude.

This thesis was completed during my research at the Chair of Dynamics and Mechatronics, Paderborn University, under the supervision of Prof. Dr.-Ing. habil. Walter Sextro. I am deeply grateful to my supervisor for believing in me, providing me with the opportunity to pursue my research, and for his continuous academic guidance throughout the years. I also thank him for serving as the first examiner of this thesis. I am grateful to the members of my examination committee for their time and support throughout the examination process. My gratitude goes to Prof. Dr.-Ing. Gunter Kullmer for serving as the second examiner of this thesis. I also extend my appreciation to Prof. Dr. Iryna Mozgova for chairing my examination committee and to Dr.-Ing. Tobias Hemsel for his academic and professional support, as well as for serving as a member of the committee.

I would like to express my heartfelt gratitude to my colleagues at the Chair of Dynamics and Mechatronics. I am especially grateful to Dr.-Ing. Amelie Bender, Lars Muth and Alexander Löwen for the successful PHM Data Challenges and our many fruitful discussions. I also wish to acknowledge Eduard Kubi and the late Martin Liekenbroecker for their invaluable assistance during the test rig development. My appreciation also goes to Marina Kassühlke and Sandra Siekmann for their kind and reliable secretarial support. I am equally thankful to all the students who assisted me in the laboratory and to those who entrusted me with supervising their theses. My sincere thanks also go to Alireza Javanmardi for the excellent academic cooperation and collaboration. I am particularly grateful to Dr.-Ing. Jan Schütte for being such an understanding office mate and for the many enriching academic discussions as well as the enjoyable conversations beyond work. I would also like to express my sincere appreciation to Dr.-Ing. James Kuria Kimotho for introducing me to the field of diagnostics and prognostics, which laid the foundation for much of my research.

Above all, I give thanks to God for His grace, strength, and guidance throughout this journey. My deepest gratitude goes to my family for their unwavering love and support. I am especially grateful to my parents, who instilled in me the value of education and continually encouraged me to pursue my goals. Finally, to my beloved wife, Precious, and our wonderful children, Éfomó and Adamosa, thank you for your patience, perseverance, unconditional love, and constant emotional support throughout my PhD journey. This achievement is as much yours as it is mine.

Enger, July 2026

Osarenren Kennedy AimiyeKagbon

Abstract

Time-varying operating conditions can obscure condition monitoring signals, either causing changes in operating conditions to be misclassified as faults, which leads to false alarms, or causing actual faults to be unidentifiable due to the influence of operating conditions on the condition monitoring signals. False alarms typically result in economic losses, while undetected faults can have serious environmental and human consequences. To enable reliable fault detection and accurate prediction of the remaining useful life (RUL) for technical systems operating under time-varying operating conditions, this work provides a comprehensive framework comprising five distinct strategies to leverage information about such time-varying operating conditions. The proposed methodology is evaluated using experimental data from accelerated run-to-failure tests on piezoelectric bending actuators and ball bearings subjected to time-varying operating conditions. Piezoelectric bending actuators are electromechanical components utilized in numerous household, medical and industrial production applications, whereas ball bearings are widely used mechanical components within more complex technical systems. The evaluation results from both application examples demonstrate that incorporating the operating conditions, for instance either via condition-invariant features or by including the operating conditions as additional inputs to machine learning models, enhances the effectiveness of fault detection and prognostics methods. Although evaluated on only two application examples, this study contributes to advancing the field of predictive maintenance, particularly for technical systems operating under time-varying operating conditions.

Kurzfassung

Zeitlich variierende Betriebsbedingungen können die Genauigkeit der Zustandsüberwachung beeinträchtigen, indem sie entweder zu Fehlklassifikationen von Betriebsbedingungsänderungen als Fehler führen, was Falschalarme zur Folge hat, oder dazu, dass tatsächliche Fehler nicht erkannt werden. Falschalarme führen typischerweise zu erheblichen wirtschaftlichen Verlusten, während unerkannte Fehler schwerwiegende Folgen für die Umwelt und die Menschen haben können. Um eine zuverlässige Fehlererkennung und eine genaue Vorhersage der Restlebensdauer technischer Systeme unter zeitlich variierenden Betriebsbedingungen zu ermöglichen, stellt diese Arbeit ein Framework mit fünf verschiedenen Ansätzen zur Nutzung von Informationen über solche zeitlich variierenden Betriebsbedingungen vor. Als Anwendungsbeispiele werden experimentelle Daten aus beschleunigten Lebensdauertests an piezoelektrischen Biegewandlern sowie Rillenkugellagern herangezogen, die jeweils zeitlich variierenden Betriebsbedingungen ausgesetzt sind. Piezoelektrische Biegewandler stellen elektromechanische Komponenten dar, die in vielen Anwendungen eingesetzt werden, während Rillenkugellager weitverbreitete mechanische Bauteile in komplexeren technischen Systemen sind. Die Ergebnisse zeigen, dass die Berücksichtigung der Betriebsbedingungen, entweder durch Merkmale, die unabhängig von den Betriebsbedingungen sind, oder durch die Einbindung der Betriebsbedingungen als zusätzliche Eingangsdaten in maschinelle Lernverfahren, die Vorhersagegüte von Fehlererkennungs- und Prognosemethoden verbessert. Obwohl nur an zwei Anwendungsbeispielen evaluiert, liefert diese Arbeit einen Beitrag zur Weiterentwicklung der prädiktiven Instandhaltung, insbesondere für technische Systeme, die unter zeitlich variierenden Betriebsbedingungen betrieben werden.

Prior Publications

Parts of this thesis have been previously published in the following articles.

Peer-Reviewed Journal Articles

AIMIYEKAGBON, O. K.; BENDER, A.; HEMSEL, T.; SEXTRO, W.: Diagnostics of Piezoelectric Bending Actuators Subjected to Varying Operating Conditions. *Electronics*, Volume 13, Issue 3, 2024

Peer-Reviewed Articles in Conference Proceedings

AIMIYEKAGBON, O. K.; BENDER, A.; SEXTRO, W.: Evaluation of time series forecasting approaches for the reliable crack length prediction of riveted aluminium plates given insufficient data. In: *PHM Society European Conference*, Volume 5, Issue 1, 2020

AIMIYEKAGBON, O. K.; MUTH, L.; WOHLLEBEN, M.; BENDER, A.; SEXTRO, W.: Rule-based Diagnostics of a Production Line. In: *PHM Society European Conference*. 2021

AIMIYEKAGBON, O. K.; LOWEN, A.; BENDER, A.; MUTH, L.; SEXTRO, W.: Expert-informed hierarchical diagnostics of multiple fault modes of a spacecraft propulsion system. In: *PHM Society Asia-Pacific Conference*, Volume 4, Issue 1, 2023

BENDER, A.; AIMIYEKAGBON, O. K.; SEXTRO, W.: Diagnostics and Prognostics for Retrofitted Systems: A Comprehensive Approach for Enhanced System Health Assessment. In: *IEEE Prognostics and System Health Management Conference (PHM)*, 2024, p. 159-164

JAVANMARDI, A.; AIMIYEKAGBON, O. K.; BENDER, A.; KIMOTH, J. K.; SEXTRO, W. HÜLLERMEIER, E.: Remaining Useful Lifetime Estimation of Bearings Operating under Time-Varying Conditions. In: *Proceedings of the 8th European Conference of the Prognostics and Health Management Society*, 2024

AIMIYEKAGBON, O. K.; LOWEN, A.; HANSELLE, R.; RIEF, T.; BECK, M.; SEXTRO, W.: Multilevel fault diagnostics for railway applications using limited historical data. In: *PHM Society Asia-Pacific Conference*, Volume 5, Issue 1, 2025

Non Peer-Reviewed Contributions and Published Dataset

AIMIYEKAGBON, O. K.; BENDER, A.; SEXTRO, W.: On the applicability of time series features as health indicators for technical systems operating under varying conditions. In: *Proceedings of the Seventeenth International Conference on Condition Monitoring and Asset Management*, 2021

AIMIYEKAGBON, O. K.; BENDER, A.; SEXTRO, W.: Extraktion und Selektion geeigneter Merkmale für die Restlebensdauerprognose von technischen Systemen trotz aleatorischen Unsicherheiten. In: *3. VDI-Fachtagung Schwingungen*. Würzburg: VDI-Berichte 2391, 2021, p. 197–210

AIMIYEKAGBON, O. K.: Run-to-failure data set of ball bearings subjected to time-varying load and speed conditions. *Zenodo*, Version 2023-04-02, 2024

Contents

Nomenclature	xiii
1 Introduction	1
1.1 Research objectives	2
1.2 Thesis outline	3
2 State-of-the-art	5
2.1 Technical understanding	7
2.2 Data understanding	8
2.3 Data preparation	9
2.4 Diagnostics	10
2.5 Prognostics	15
2.6 Diagnostics and prognostics evaluation methods	18
3 Methodology	23
3.1 Consideration in the pre-data acquisition phase	23
3.2 Consideration in the post-data acquisition phase	24
3.3 Framework for fault detection and prognostics under time-varying operating conditions	31
4 Fault detection and prognostics of piezoelectric bending actuators	35
4.1 Technical understanding	35
4.2 Data understanding and preparation	41
4.3 Fault detection approaches	49
4.4 Performance evaluation of the fault detection approaches	57
4.5 Prognostics approaches	59
4.6 Performance evaluation of the prognostics approaches	63
5 Fault detection and prognostics of ball bearings	65
5.1 Technical understanding	66
5.2 Data understanding and preparation	82
5.3 Fault detection approaches	92
5.4 Performance evaluation of the fault detection approaches	99
5.5 Prognostics approaches	101
5.6 Performance evaluation of the prognostics approaches	109
6 Conclusion and Outlook	113
References	115
Appendix	137
A Characteristics and further results of the piezoelectric bending actuators	137

B	Characteristics and further results of the ball bearings	139
---	--	-----

Nomenclature

For brevity, the following lists include only symbols and abbreviations that appear multiple times or are of importance.

List of symbols

α	Permissible error margin
ε	Random measurement noise
$\tilde{a}_{ISO}$	Temperature-based modification factor
A_{thresh}	Vibration-based threshold value
$aRUL$	Actual remaining useful life
C	Basic dynamic radial load rating
C_0	Basic static load rating
F_{amp}	Dynamic load amplitude
F_{dyn}	Bearing dynamic load
F_{stat}	Bearing static load
f_r	Resonance frequency
f_s	Sampling frequency
$i(t)$	Output current measured over time t
$\hat{i}$	Root mean square value of the output current
L_{10}	Bearing basic rating life
n_k	Output speed at the k^{th} measurement
P	Equivalent dynamic bearing load
p	Bearing life exponent
$pRUL$	Predicted remaining useful life
RUL_k	Remaining useful life at time step k
T_{eq}	Equilibrium temperature
T_k	Temperature value at time step k
T_{thresh}	Temperature-based threshold value
t	Time
t_{EoL}	Time at end-of-life
t_k	Current time step k

$u(t)$	Input voltage measured over time t
$\hat{u}$	Root mean square value of the input voltage
$v(t)$	Output velocity measured over time t
$\hat{v}$	Root mean square value of the output velocity

Abbreviations

BoL	Beginning-of-Life
C-MAPSS	Commercial Modular Aero-Propulsion System Simulation
CRISP-DM	Cross Industry Standard Process for Data Mining
DAQ	Data Acquisition
DMME	Data Mining Methodology for Engineering applications
EEG	Electroencephalogram
EMG	Electromyography
EoL	End-of-Life
FMEA	Failure Modes and Effects Analysis
GPR	Gaussian Process Regression
HCTSA	Highly Comparative Time-Series Analysis
KDD	Knowledge Discovery in Databases
KDDM	Knowledge Discovery and Data Mining
nRMSE	Normalized Root Mean Squared Error
PHM	Prognostics and Health Management
RMS	Root Mean Square
RUL	Remaining Useful Life
SEMMA	Sample, Explore, Modify, Model, and Assess
SHM	Structural and Health Monitoring
SVM	Support Vector Machine
wMAPE	Weighted Mean Absolute Percentage Error

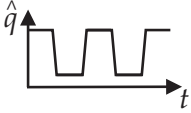
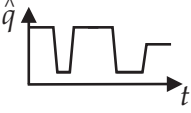
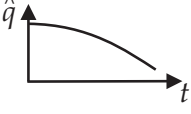
1 Introduction

According to Isermann and Ballé [1], fault detection is the determination of the impermissible deviation from the standard, usual or acceptable condition of at least a characteristic feature of a system via condition monitoring signals, while prognostics is defined as the process of estimating the RUL of a technical system degrading either due to normal operation with no fault symptoms or due to normal operation with a detected fault [2].

Cyber-physical systems autonomously change their behavior to respond to dynamic environments through self-adaptation [3]. Mechatronic systems are subjected to time-varying conditions, such as exemplarily presented in Table 1.1 for an arbitrary condition. Time-varying conditions can either be set by the operator or can be due to the environment. The profile of time-varying conditions at a given time instant t_k could be constant, periodic, with a given amplitude and frequency or stochastic as classified in [4], [5]. In summary, mechatronic and cyber-physical systems are complex systems that can operate under time-varying conditions. For example, wind turbines are typically subjected to a combination of varying loads, such as, gravitational load, centrifugal load and wind load [6]. In an aircraft, flight speed and altitudes vary between takeoff and landing and the weather conditions could also vary during a flight [7]. Lithium-ion batteries, such as in electric vehicles and photovoltaic systems, are often subjected to varying environmental conditions, such as temperature variations [8]. Where different and customized products are produced to meet specific customers' requirements, computer numerical control (CNC) machines of a production industry are typically subjected to varying operating conditions because each product might vary in composition, material and dimension [9]. Such time-varying conditions pose several challenges in fault detection and prognostics, as a result of two influencing factors. Firstly, sudden jumps in the operating conditions, such as shocks can expedite fault progression or even lead to sudden failure [10]. Secondly, because time-varying operating conditions significantly influence the degradation process and consequently the condition monitoring signals [10], either faults might not be detected or changes in the operating conditions could be misclassified as faults and consequently result in false alarms [11]. On the one hand, undetected faults diminishes product quality, faults progressed into failure lead to long-term downtime, huge economic losses, and devastating human and environmental consequences for safety-critical systems. On the other hand, false alarms lead to unnecessary maintenance planning and consequently to increased economic cost. Thus, to accurately predict developing faults or failure before occurrence or detect developed faults at an early stage, the consideration of the underlying operating conditions is inevitable.

In the following, the term *system* is used interchangeably to imply a component, a subsystem or a system of subsystems. However, this work builds on the assumption that the failure of a critical non-redundant component leads to the failure of the total system. Thus, this work focuses on the component-level.

Table 1.1: Exemplary categories of time-varying conditions adapted from [4], [5]

Category	Profile	Typical example
Stepped cyclic		Repetitive tasks of a robot with two or more steps, such as pick and place.
Stepped non-cyclic		Non-repetitive tasks with two or more steps, such as required in small batch or one-off production.
Monotonic		Machine run-up or run-down; Flight takeoff or descent.
Transient		Short-term events, such as shocks, and surges.
Random		Wind load; Earthquake on structures, such as wind turbines.

1.1 Research objectives

Time-varying operating conditions can accelerate the degradation rate and simultaneously obscure condition monitoring signals. This may either lead to faults becoming undetectable, eventually resulting in system failure, or to variations in operating conditions being misclassified as faults, causing false alarms. Therefore, understanding the relationship between time-varying operating conditions, the degradation process, and condition monitoring signals is crucial for improving the accuracy of fault detection and prognostic models. Enhanced model accuracy helps reduce unplanned downtime, maintain product quality, and prevent adverse economic, human, and environmental consequences, especially in safety-critical systems.

In developing fault detection and prognostic models, the extraction and selection of degradation-relevant features constitute a vital step, as raw condition monitoring signals often do not directly reveal underlying degradation trends. Moreover, since time-varying operating conditions, which can mask condition monitoring signals and the features derived thereof, are not always (synchronously) measured or even measurable, it is important to investigate features that reliably reflect degradation regardless of operating condition variations.

The effects of time-varying operating conditions can be addressed in multiple ways during the development of fault detection and prognostic models, each offering different degrees of effectiveness. It is therefore essential to explore various approaches and identify the most practical and promising approaches for early fault detection and prognostics under

these time-varying operating conditions.

The main research objective is to provide a comprehensive framework for reliable fault detection and prognostics in technical systems subjected to time-varying operating conditions. Thus, from the preceding paragraphs, the specific research objectives can be summarized as follows:

- Identify possible relationships between the underlying operating conditions, the degradation process, and the condition monitoring signals.
- Determine suitable features that are more sensitive to degradation than to changes in operating conditions.
- Investigate methods and assess the benefits of incorporating operating condition information into fault detection and prognostic tasks.

1.2 Thesis outline

The thesis is organized into six chapters. The introductory chapter begins by defining fault detection and prognostics, establishing the need for research focused on technical systems operating under time-varying conditions. Additionally, examples of time-varying operating conditions are presented, and the chapter concludes with a summary of the research objectives.

Chapter 2 reviews and compares various knowledge discovery and data mining (KDDM) processes, with a primary focus on prognostics and health management (PHM). To address recent advancements in machine learning, the traditional PHM cycle is extended by integrating steps derived from engineering and industry-relevant KDDM frameworks. The phases of the extended PHM cycles, namely, Technical Understanding, Data Understanding, Data Preparation, Diagnostics, Prognostics, Evaluation Methods, and Decision Management, are then described in detail.

Chapter 3 explores multiple approaches to account for time-varying operating conditions, beginning with approaches applicable prior to data acquisition, particularly during the technical understanding phase. Following data acquisition, five approaches for incorporating time-varying conditions are discussed: specialized signal analysis techniques, condition-invariant feature engineering, condition-centered feature scaling, condition-specific feature standardization, and embedding operating conditions within the modeling phase. The chapter concludes with a flowchart that systematically presents these methodologies, forming a framework for fault detection and prognostics under time-varying operating conditions.

Chapters 4 and 5 focus on two application examples to evaluate the proposed methodologies, namely piezoelectric bending actuators and ball bearings, respectively. Each chapter starts with an overview of the application. Under the *Technical Understanding* section, a failure modes and effects analysis (FMEA), potential monitoring techniques, test rig development, experimental setup and measurement procedure are described. The *Data Understanding and Preparation* section investigates how operating conditions influence the degradation process and condition monitoring signals. Subsequent sections present fault detection and prognostic approaches, while examining the five aforementioned approaches and employing machine learning methods. Both chapters conclude with an evaluation of the model performances.

Chapter 6 summarizes the key research findings, reflecting on the original research objectives. It also discusses potential future research directions aimed at further improving the accuracy of fault detection and prognostics models, particularly in the context of time-varying operating conditions.

2 State-of-the-art framework for diagnostics and prognostics

This chapter offers a theoretical background by describing the key steps in developing diagnostics and prognostics models. Moreover, it reviews and compares diverse KDDM methodologies, focusing primarily on PHM.

The workflow that links condition monitoring signals, fault detection and prognostics can be described under the general KDDM process, which involves the application of data mining methods to seek knowledge from data via data analysis and pattern extraction [12]. There exist several KDDM models, such as knowledge discovery in databases (KDD) [12], [13], sample, explore, modify, model, and assess (SEMMA) [12], cross industry standard process for data mining (CRISP-DM) and several extensions of CRISP-DM [12], [14], such as data mining methodology for engineering applications (DMME) [15]. According to Fayyad et al. [13], the uttermost goal of KDD is to extract high-level knowledge from low-level data in databases. SEMMA was defined by and incorporated in the commercial software SAS[®] Enterprise Miner[™] of the SAS Institute [12]. CRISP-DM is an industry-oriented and application-/tool-neutral model proposed in 1996 by a consortium of four companies: Daimler-Benz (now Mercedes-Benz Group AG, one of the world's leading automotive companies), Integral Solutions Ltd. (now IBM, a commercial data mining solutions provider), NCR (a database provider), and OHRA (a notable Dutch insurance company) [12], [14]. DMME is an holistic extension to CRISP-DM incorporating three new steps for engineering applications, namely technical understanding, technical realization and technical implementation [15]. A detailed literature survey of the KDDM models, adaptations to CRISP-DM and related process models can be found in [12], [16], [17].

As opposed to models, which provide a general set of steps for KDDM, the four-part ISO 13374 standard provides a functional specification for condition monitoring and diagnostics [18], [20], [21], [22]. A notable implementation of the ISO 13374 standard is the MIMOSA open system architecture for condition based monitoring (OSA-CBM) [23]. Finally, in the PHM society, the PHM cycle is established [19], [24]. In Table 2.1, the different steps of major KDDM models are compared against the ISO 13374 standard and PHM cycle. As can be deduced from Table 2.1, KDD has the most detailed steps and begins with understanding the application domain and identifying the goal of the KDDM process [13]. CRISP-DM is similar to KDD, but is more industry-oriented. Except for the first step and explicit evaluation step, the other steps of KDD and CRISP-DM can be found represented in the ISO 13374 standard and PHM cycle. It can be presumed that the PHM cycle is based on the ISO 13374 standard because of the similarity between the steps. However, as opposed to the ISO 13374 standard which defines a strict top to bottom data/information flow, the PHM cycle allows for a feedback loop via the cyclic nature. Furthermore, depending on the task at hand, diagnostics and/or prognostics can be performed. Nevertheless, the PHM cycle assumes that a condition monitoring system is already set and hence begins with data acquisition. Because this is not always the case

Table 2.1: Comparison of three major KDDM methods to ISO 13374 and PHM cycle adapted from [12]

Method Step	KDD [12], [13]	SEMMA [12]	CRISP-DM [12], [14]	ISO 13374 [18]	PHM cycle [19]
1	Application domain understanding	-	Business understanding	-	-
2	Target dataset creation	Sample Explore	Data understanding	Data acquisition	Data acquisition
3	Data preprocessing	-	Data preparation	Data manipulation	Data preprocessing
4	Data transformation	Modify			Feature extraction
5	Task selection	-	Modeling	State detection Health assessment Prognostic assessment	Enhanced diagnostics Prognostics
6	Algorithm selection	Model			
7	Data mining				
8	Interpretation Evaluation	Assess	Evaluation	Advisory generation	Decision management
9	Act on discovered knowledge	-	Deployment		

and to allow for more flexibility, the PHM cycle is supplemented with steps from KDDM models in this thesis, as depicted in Figure 2.1. In the following, each step of Figure 2.1, except *decision management*, will be elaborated on, while reviewing relevant literature. The *decision management* step is beyond the scope of this thesis, hence it is not further considered here.

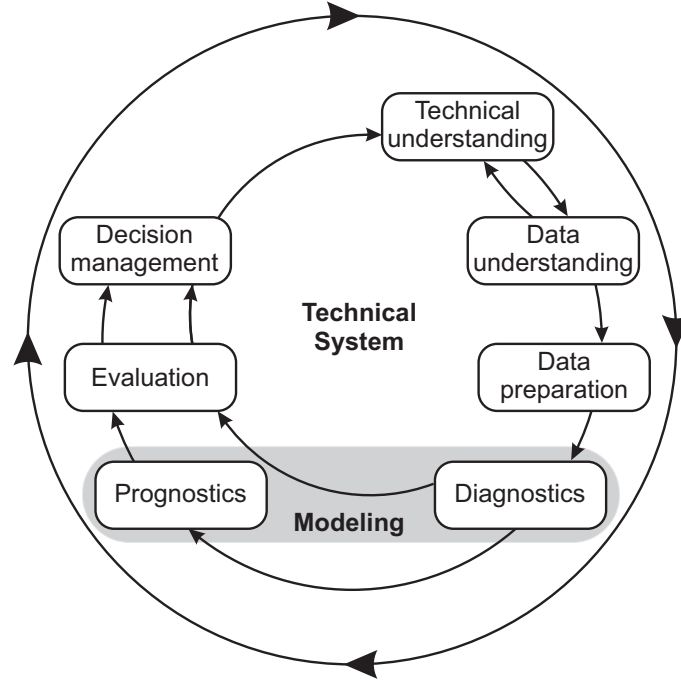


Figure 2.1: PHM cycle [19] extended with steps and inspiration from CRISP-DM [14] and DMME [15]

2.1 Technical understanding

Before instrumenting a condition monitoring system on a technical system, due to initial costs, it is important to conduct an in-depth survey of the applicability and potential of condition monitoring. Thus, the *Technical understanding* step, which helps to uncover the benefits of a condition monitoring system is incorporated in the traditional PHM cycle from DMME [15]. According to Huber et al. [15], this step involves an assessment of the available system information from different sources, such as expert information about the system, the process and related parameters. Furthermore, this step entails the definition of relevant measurable physical parameters and their influence on the system or process. Concepts for measuring the physical parameters and an experimental plan, for example, through design of experiments are included in this step. Because structural and health monitoring (SHM) for structures is closely related to condition monitoring of technical systems, SHM methods are also applicable here [25]. However, the ISO 17359 [26] provides a general procedure for condition monitoring and diagnostics, therefore it can also be followed. The parts of the ISO 17359 relevant for this step are represented in the flowchart as illustrated in Figure 2.2. The ISO 17359 begins with a cost benefit analysis under consideration of life cycle cost and consequences of unplanned downtime. It is followed by a survey of the technical system to be monitored, interacting systems and processes. Furthermore, the input and output parameters and existing monitoring

system, if available, are evaluated. Subsequently, a reliability and criticality analysis is performed to identify critical components, possible failure modes and their influence qualitatively via FMEA or quantitatively via failure mode, effect, and criticality analysis [26]. If the identified fault/failure modes are measurable via physical parameters, then the appropriate measurement techniques are selected. Finally, the measurement setup, including the measurement interval, sampling frequency, sensor placement, measurement range, sensitivity and applicability of the selected measurement techniques should be considered.

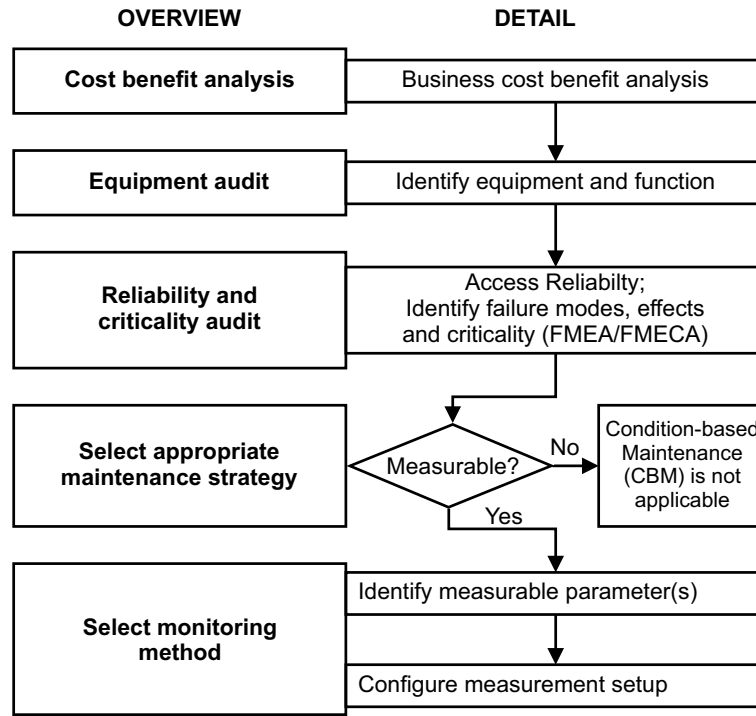


Figure 2.2: Excerpt from the ISO 17359 standard, relevant to the technical understanding step

2.2 Data understanding

This step is drawn from CRISP-DM and consists majorly of initial data acquisition, data description, data exploration and data quality verification [14]. Based on the selected measurement techniques and sensor configurations, the measurable input and output physical parameters are acquired. Other information, such as planned interruption of operation, system configuration change or component change should also be documented [26]. According to ISO 17359 [26], a baseline data should be acquired typically after a wear-in period, where the system is in stable and normal state. When the system is operating under varying operating conditions, it is suggested to acquire baseline data at the distinct operating conditions. However, it might be impossible to establish baseline data at every distinct operating condition, where the future operating condition profile is unknown a priori. Furthermore, a baseline at distinct operating conditions might also be infeasible, where the transition between two operating conditions is continuous and not easily distinguishable. After the initial data acquisition, the data is examined if it satisfies relevant requirements, such as if the data is in the appropriate and required format [14]. Furthermore, an initial criteria for the identification of a fault occurrence can also be set based

on the initial data [26]. Also relevant in this phase is the data exploration, for example via different data visualization techniques [14]. The relationship between the measured parameters and characteristics of the data, such as trends can be uncovered through visualization. Furthermore data quality issues, such as missing values can also be detected via visualization. According to ISO 17359 [26], poor data quality can result from factors, such as sensor fault, improper sensor mounting and inappropriate measurement configuration. A review of different types of sensor faults and sensor fault detection methods can be found in the article by Li et al. [27]. A notable sensor fault detection method from their review article is redundancy, that is, either via hardware or analytical redundancy [27]. Hardware redundancy involves the measurement of the same parameter with multiple sensors, because it is highly unlikely that the multiple sensors would simultaneously be faulty [27]. Analytical redundancy is a cost- and space-efficient alternative which builds on the generation of a fault-free sensor estimation, for example through a mathematical model. A sensor fault is then detected by comparison to the redundant physical sensor(s) or estimated measurements [27], [28].

2.3 Data preparation

The data processing and feature extraction phases of the classical PHM cycle is replaced by a generic data preparation phase of CRISP-DM to address developing new trends, such as the emergence of deep learning methods, where data is employed directly as input for modeling without prior manual feature extraction. Even in the application of deep learning methods, the input data has to be prepared in the right format. In general, data preparation involves data selection, data cleaning, data construction, data integration and data formatting [14]. From the preceding phase of data understanding, data selection pertains to the exclusion of specific data, for example if they are redundant or depending on predetermined exclusion criteria [14]. Data cleaning typically involves outlier removal and noise filtering. Outlier can generally be detected via clustering-, density-, statistical-, and distance-based techniques [29]. Noise is often unavoidable in sensor measurements because of either electrical interference or mechanical induced vibrations. The electromagnetic field of the power supply results in the line frequency of 50 Hz and harmonics thereof. Vibration of neighboring systems, vibration of electronic components and cables in the measurement path also result in noise. Depending on the amplitude of the noise, the sensor measurements might be unusable [30]. Thus, noise is generally an unwanted component of a signal. Although noise might not completely be eliminated, there exists several techniques to attenuate the noise portion of a signal, while maintaining the wanted portion of the signal. Some noise filtering techniques are infinite impulse response (IIR) digital filter, phase space filters, Wiener filters, wavelet denoising, empirical mode decomposition (EMD) and adaptive noise canceling [30], [31], [32]. After the data cleaning process, if necessary, the next phase is data construction. As the name implies, data construction involves data transformation as in dimensionality reduction or feature extraction to derive new informative data [13], [14]. Notable dimensionality reduction techniques are principal components analysis (PCA), local linear embedding, Sammon mapping and multilayer autoencoders [33]. Features can typically be extracted in the time-, frequency- and time-frequency domain. Time domain features include statistical features, time derivatives, or other mathematical operations in the time domain. After transformation in the frequency domain, for example via the fast Fourier transform (FFT), statistical features

can also be computed in the frequency domain. Other frequency domain features include maximum power of the power spectral density (PSD), the amplitude at characteristic frequency components, such as bearing fault frequencies or gear mesh frequency and sum of amplitudes within a bandwidth. Time-frequency transformation techniques include short-time Fourier transform, Wigner-Ville distribution, discrete and continuous wavelet transform, synchro-extracting transform, synchrosqueezed wavelet transform and Stockwell transform [34], [35]. Because condition monitoring signals are often time series data, several toolboxes, such as the highly comparative time-series analysis (HCTSA) [36], [37], electromyography (EMG) and electroencephalogram (EEG) feature extraction toolbox [38], [39], time series feature extraction on basis of scalable hypothesis tests (tsfresh) [40], and flexible time series processing & feature extraction (tsflex) [41] can assist in the feature extraction process. After feature extraction, a subset of relevant features can be selected as health indicators either via filter, wrappers or embedded methods [42]. The final steps in the data preparation process involve integrating the health indicators from multiple sources and putting the data in the suitable format for the desired models. Here, an health indicator is a feature that allows the quantification of a system's degradation level derived from condition monitoring signals.

2.4 Diagnostics

Fault or failure diagnostics generally involves *detection*, *isolation* and *identification* [1]. According to Isermann and Ballé [1], failure occurs when a system is permanently unable to perform a required function when subjected to specified operating conditions [1]. From the preceding definitions, a fault detected early allows for enough lead time before failure. Furthermore, since failure is apparent, the detection, isolation and identification of fault will be focused on here. Fault detection involves a binary decision if there is a fault or not. Fault isolation involves the determination of the location, kind and time of the detected fault. Fault identification involves the determination of the time-variant behavior and size of the isolated fault [1]. Depending on the time-dependency of the fault, faults can be classified as abrupt, intermittent or incipient [43]. An abrupt fault occurs suddenly and stepwise, an intermittent fault exhibits symptoms at irregular intervals, while an incipient fault develops slowly with gradual symptoms, resulting in a blurred or fuzzy boundary between healthy and faulty behavior [43], [44].

Depending on the effect of the fault on the process or system, fault can also be additive or multiplicative. Additive fault generally influences the sensor output by an offset via superposition. Contrarily, multiplicative fault influences the internal system parameters, such as the resistance, stiffness coefficients, damping coefficients or inertia constants [45]. Technical systems and specifically machine elements, such as gears and bearings often possess microstructural flaws, such as inclusions and asperities inherent of imperfection in the production process, which is a source of microcracks. Due to the time-dependent loading situation, especially in the contact zone of rolling or sliding materials, such initial microcracks gradually propagate into larger cracks as a result high stress in the vicinity of the microcracks [46]. Such propagated cracks consequently influence the internal material parameters, such as the damping or stiffness properties. Thus, technical systems are typically subjected to incipient and multiplicative faults. Based on the articles [47], [48], [49], [50], [51], [52], [53], methods for fault diagnostics can be broadly classified into model-based, history-based and hybrid method, as depicted in Figure 2.3.

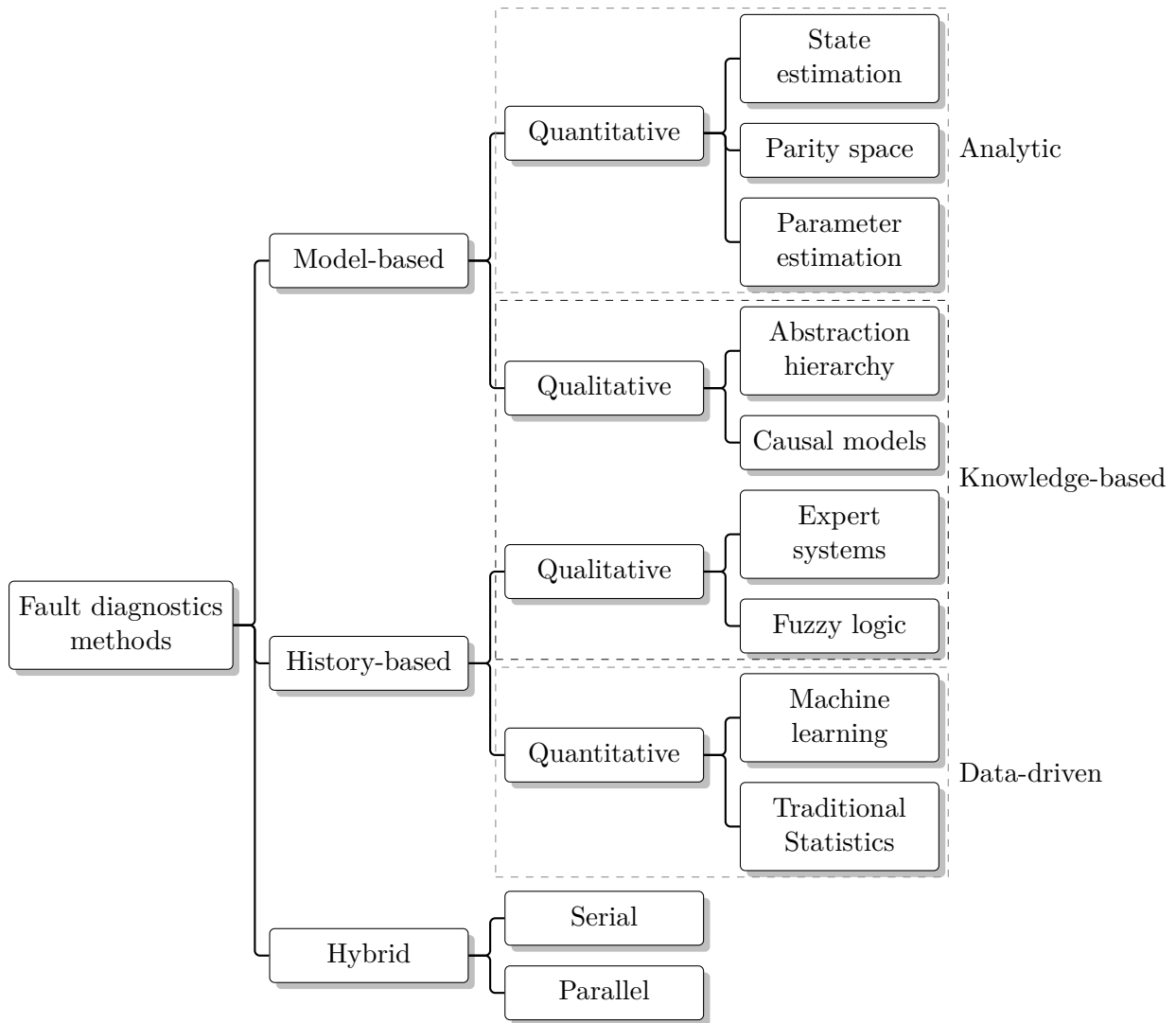


Figure 2.3: Classification of fault diagnostics methods based on several articles [47], [48], [49], [50], [51], [52], [53], [54]

Although some historical data is required in model-based methods, for example to estimate model parameters, and some model is derived in history-based methods, the fundamental approaches to fault diagnostics differs between the two. Model-based method typically builds on founded domain knowledge of the system, while history-based method relies on measured input/output history of the system. Depending on the proportion of domain knowledge to process history required in model-based and history-based methods, both methods can be further classified in quantitative and qualitative methods. The proportion of domain knowledge to process/system history is higher in qualitative methods, as opposed to quantitative methods, where the reverse is observed. Thus, on the one hand, qualitative methods are also known as knowledge-based methods. On the other hand, quantitative history-based methods are often referred to as data-driven methods, while quantitative model-based methods are often referred to as analytical methods. Although hybrid methods can also refer to the combination of either quantitative and qualitative model-based or history-based methods [49], hybrid methods are described here as the combination of model-based and history-based methods, while exploiting the advantages of both methods. Based on the level of dependency of the combined methods, hybrid

methods can broadly be classified in serial and parallel approach. In a serial approach and according to the modeling details, either the model-based diagnostics model depends on the outputs of the history-based model or the history-based model is dependent on the outputs of the model-based model. However, the serial approach, where the output of the history-based model is fed as input to the model-based model is more likely in cases where certain dynamics are not accurately described by the model-based model [54]. As opposed to the serial approach, both models are deployed simultaneously and their outputs are combined as an ensemble in the parallel approach [54]. In the following paragraphs, common model-based and history-based methods will be briefly introduced, because hybrid models builds on both models.

2.4.1 Model-based methods

As depicted in Figure 2.3, model-based methods can broadly be classified in quantitative and qualitative methods. Commonly employed quantitative model-based methods are state estimation, parameter estimation and parity space. Furthermore, abstraction hierarchy and causal models are commonly employed qualitative model-based methods.

Observers and Kalman filters are two widely employed methods for deterministic and stochastic state estimation, respectively. Observers generally generate a set of residuals for fault diagnostics through the estimation of the system output derived from partial or complete input/output measurements [47], [55], [56]. Observers are typically modeled as state-space equations and fault-specific, in the sense that a single observer is developed per fault mode and therefore is only sensitive to that particular fault mode [47]. Different from observers that consider estimation error as a residual, Kalman filters and extensions thereof consider innovation. According to Mehra and Peschon [57], innovation is the difference between the actual system output and the expected output based on the model and the previous output data, that is, at time prior to the current time. On the contrary, residual considers output data up to the current time.

A similar, but simpler approach with comparable results to observers is the parity space [56]. The parity space method, which is also known as parity equations or parity relations builds on the parity or consistency comparison of the system model to the measured system output. In the parity space, a fault is detected if the residual between the system model and the measured system output exceeds a predetermined threshold [47], [55], [56]. According to Isermann [56], parity equations are better suited for the detection and isolation of additive faults, while parameter estimation methods are suitable for detection of multiplicative faults. Since system parameters often depend on the physical coefficients, such as stiffness and damping coefficients, parameter estimation methods determine the fault-induced changes in such parameters via input/output measurements and a dynamic model of the system [56].

As the name suggests, abstraction hierarchy allows for a general representation of a system's functionality via input/output relationship of subsystems or entities of the system in an hierarchical manner. Groups of subsystems and their functions are described at the top of the hierarchy, while individual entities, such as components or subsystems and their functions are described at the bottom of the hierarchy. Fault is diagnosed top-down, where a fault at the higher level is traced down to fault of one or more components at the bottom level [48]. The relationship between cause (fault) and effect (symptoms) can be modeled in causal relations of the form: fault $\rightarrow$ events $\rightarrow$ symptoms. Simply put, in the case of a fault, possible intermediate events and functional changes are identified

[49], [56], [58]. Causal models or relations are often represented graphically, such as in signed directed graphs and fault trees [49], [56], [58]. Signed directed graph or signed digraph is a graph with nodes and directed positive or negative weighted edges (arcs). The nodes in the graph represent events, states or variables and the edges represent the relationship between the nodes. The weighted signs in the edges represent the direction of change, as in positive or negative influence [49], [56], [58]. As opposed to signed digraphs that mainly consist of OR logic nodes, different logic nodes, namely OR, AND and XOR can be used in fault trees [49]. In fault trees, the conditional probability of events and symptoms are employed to determine the probability of a fault occurring, by modeling the propagation of a fault to its manifestations (symptoms) via intermediate events [44]. Although, fault tree is widely employed in the analysis of system reliability and safety, it has limited application to fault diagnostics of technical systems that undergo incipient, that is, slowly developing faults [44], [49], [56].

2.4.2 History-based methods

As the name suggests, history-based methods are established on process history, for example in the form of measured input/output parameters, fault/failure logs, which incorporate fault/failure times and performed maintenance/repair actions. As depicted in Figure 2.3, history-based methods can broadly be classified in qualitative methods and quantitative methods. In the following, some commonly employed qualitative and quantitative methods are briefly described.

Expert systems and fuzzy logic both derive qualitative knowledge from quantitative process history. Domain experts have vast knowledge in their field of expertise. Thus, in expert systems, their knowledge together with other process history information, such as operational settings, degree of use/aging and fault statistics are collected in so-called knowledge base [49], [50], [55]. Based on the acquired knowledge, **if-then-else** rules are derived and inference procedures for diagnostics are developed. A further component of expert systems is an interpretation module, which explains the derivation of diagnostic conclusions and the reason behind the conclusions, such as the cause, type and size of the fault [49], [55]. Expert systems are typically software applications dedicated for a specific system and known fault modes [49], [50].

As previously stated, fault in technical systems is incipient and thus the transition from healthy to faulty is often fuzzy. Instead of a binary decision of 0 (healthy) and 1 (faulty), fuzzy logic allows values in the range $[0, 1]$ and thus allows some degree of healthiness and faultiness. Furthermore, fuzzy logic allows the representation of uncertainties and inaccuracies [59], [60]. Fuzzy logic are often employed as an extension to other knowledge-based methods, such as in signed digraphs, fault trees and expert systems [44], [48], [49], [50].

Data-driven techniques such as traditional statistics and machine learning are two frequently employed quantitative history-based methods. A system is subject to random fluctuations, even under normal operation. When abnormal events, such as a fault occurs, the underlying distributions or parameters thereof might change. Statistical methods, such as Bayes classifier employed for fault classification, are based on the knowledge of a priori class distributions. Furthermore, based on the assumption that a process will remain in a state of statistical control under normal operation, univariate statistical methods, such as statistical process control charts are also commonly employed in process quality control and fault detection [49]. Machine learning is a sub-field of artificial

intelligence, where knowledge is learned autonomously from data, similar to the human learning process [61]. As presented in Figure 2.4, machine learning techniques can generally be classified in supervised, unsupervised, semi-supervised and reinforcement learning. Furthermore, deep learning can be seen as a sub-field of machine learning for big data.

In supervised learning a relation between a set of known input and labeled output data, namely a target, is learned in a model. The learned model is then employed to predict output(s) for new and probably previously unseen input data [61]. Depending on if the output data is (continuous) quantitative or categorical data [62], supervised learning can be divided in classification and regression problem. Classification involves categorical data, where the output is a class membership, that is, which class the input data belongs. Generally, fault detection can be tackled as a binary-class classification problem, while the isolation of the faulty component can be tackled as a multi-class classification problem. Fault identification typically involves the determination of the size of the isolated fault, therefore it can be considered as a regression problem. As opposed to classification, the output in regression is quantitative data and hence is discussed under prognostics in Section 2.5.

In unsupervised learning, the input data is unlabeled, hence the task involves learning underlying pattern from the input data without a specified target variable. A major application of unsupervised learning technique in anomaly or fault detection is through clustering, where the input data is partitioned into two (distinct) clusters of healthy or unhealthy instances.

Where labeled data is partially available, semi-supervised learning utilizes the available labeled data to train a model and the unlabeled data to improve the model [61]. In reinforcement learning, the data is explored to produce a set of decisions, with the goal of maximizing some notion of reward [61].

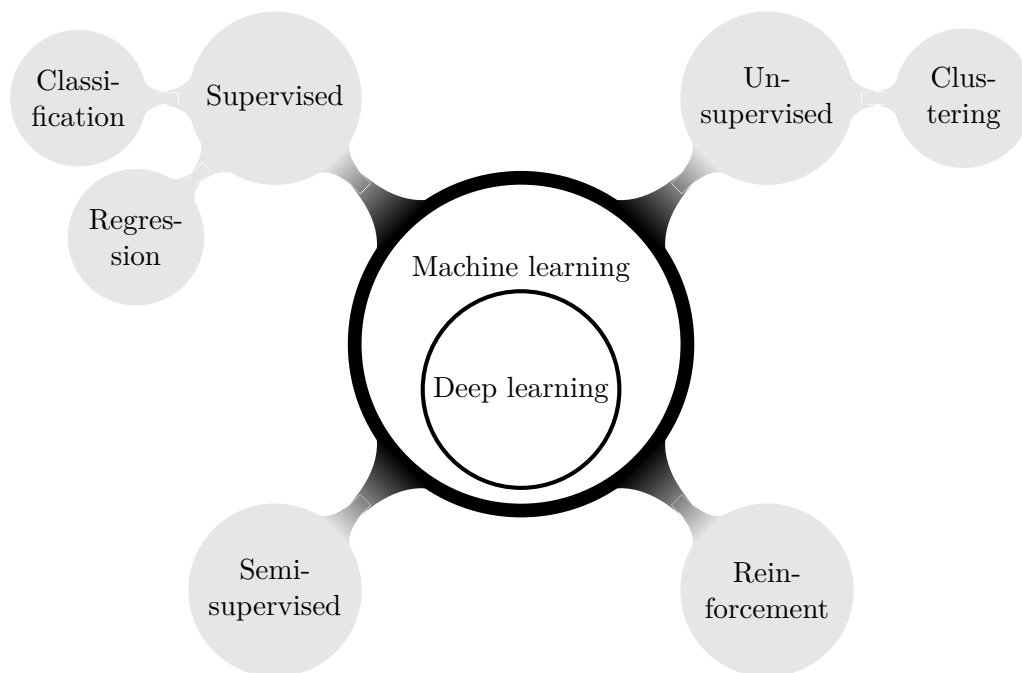


Figure 2.4: Classification of machine learning methods (adapted from Gollapudi [61])

2.5 Prognostics

Prognostics involves the estimation of the RUL. From the surveyed literature, there are several classifications of methods for prognostics. For example, most researchers classify prognostics approaches in data-driven, physics-based and hybrid approach [63], [64], [65], [66], [67]. Some distinguish prognostics approaches in physics-based, data-driven, knowledge-based and combination approach [68], [69]. Heng et al. [70] categorized prognostics methods in traditional reliability approach, data-driven, physics-based and integrated approach. Notwithstanding, these different classification attempts of prognostics methods are represented under the three broad categorization presented in Section 2.4 and as depicted in Figure 2.3, namely in model-based, history-based and hybrid approach, such that knowledge-based, traditional reliability and data-driven approaches are represented under the history-based approach, while combination and integrated approaches can be considered hybrid approaches. Since an overview of the methods have been presented in Section 2.4 for diagnostics, the commonly employed methods for prognostics are focused on here.

2.5.1 Model-based methods

Model-based or often termed physics-based methods relies on physical modeling of a system, under consideration of the degradation effects. Mechanistic models, such as Physics of failure (PoF) models or first principle models in form of differential equations or state-space models are often considered. Furthermore, Bayesian filtering techniques, such as the Kalman and particle filter and degradation models, such as the gamma and Wiener process are often employed to estimate the RUL of technical systems.

According to Gu and Pecht [71], [72], PoF modeling which have roots in reliability prediction typically incorporates loading conditions, failure modes, mechanisms and effects analysis (FMMEA) and damage models. Hence, when combined with condition monitoring signals, it links diagnostics with prognostics. In cases where the fault/failure modes are known a priori, monitored signals are employed directly in conjunction with respective damage models in prognostics. Fatigue models, such as the Coffin-Manson's strain-based model, the Basquin's stress-based model and crack growth models, such as the Paris-Erdogan equation have been widely employed in prognostics of electronic and mechanical components [63], [71], [72], [73], [74], [75], [76]. Other damage models for rotating machinery components and specifically for gears and bearings can be found in the review article by Cubillo et al. [77].

In general, under consideration of measured data up through time t , stochastic filtering involves the current state-variable estimation, while prediction involves the future state-variable estimation, that is, the state-variable at time $t' > t$ [78], [79]. Bayesian filtering typically involves the estimation of the current state under consideration of measurements up through the current time t via the probability density function (PDF) [80], [81]. The Kalman and particle filter are commonly employed Bayesian filtering techniques for the prediction of the RUL. On the one hand, the classical Kalman filter provides an exact solution for the posterior PDF under the assumption of a linear process and Gaussian white noise, while its variants, such as the extended Kalman and unscented Kalman filters are employed for non-linear processes. On the other hand, particle filters are employed for non-linear processes with non-Gaussian white noise [66], [79], [81], [82]. A non-exhaustive

list of application examples, such as in aircraft actuator system, wind turbine and lithium-ion batteries can be found in the review article by Jouin et al. [81].

The gamma and Wiener processes are both stochastic degradation modeling approaches mostly employed in reliability analysis. However, they have also found application in the prediction of the RUL. The gamma and Wiener processes are categorized under continuous-time Markov processes with independent increments, where the Markov property implies that a process only depends on the current state and therefore is memoryless [83], [84]. The gamma process is a stochastic process with independent and positive increments that follow the gamma distribution. Thus, it is mostly employed for modeling incipient monotonic degradation [83], [84]. Contrarily, the Wiener process is applicable for degradation processes that are not necessarily monotonous, such as resulting from self-healing or fluctuation in environmental or operating conditions. The Wiener process, also termed Brownian motion with drift or Gaussian process is a stochastic process with positive increments and decrements that follow the Gaussian distribution [83], [84]. An application of the Wiener process for the estimation of the RUL of lithium-ion batteries can be found in the article by Tang et al. [85] and further applications and adaptations of the Wiener process can be found in the review article by Zhang et al. [84].

2.5.2 History-based methods

History-based methods build on the process data collected over time to predict possible future trends and estimate the RUL. Commonly employed methods include trend extrapolation, time series forecasting methods and machine learning techniques.

Trend extrapolation is one of the simplest methods, where an empirical model is fitted to historical data. Under the assumption of a similar future trend, the fitted model is then employed to extrapolate or predict future progression and estimate the RUL. Typically employed models are polynomial, exponential and power models [86]. For example, Wang and Xiang [87] employed a gradient descent optimized exponential model to estimate the RUL of ball bearings.

Since sensor-based condition monitoring signals are time series data, time series forecasting methods are applicable in predicting future (fault) paths and consequently in estimating the RUL [88]. Although autoregressive model, moving average model and exponential smoothing have mostly found application in PHM-related literature as a data smoothing technique, other time series forecasting methods such as autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) models and extensions thereof have also found application in estimating the RUL [86]. As the name suggests, ARMA is a combination of an autoregressive and moving average model, while ARIMA is a generalized ARMA model applied to a differenced data [88]. A notable application example of time series forecasting methods is in the prediction of the RUL of electrochemical cells. For example, Long et al. [89] employed a particle swarm optimized autoregressive model to predict the RUL of lithium-ion batteries. Also to predict the RUL of lithium-ion batteries, Zhou et al. [90] utilized an ARIMA model after decomposing the data by the EMD.

Regression-based methods are the most commonly employed machine learning techniques in prognostics. In terms of prognostics, the input data are raw condition monitoring signals or features derived thereof, while the output data is often the RUL. Furthermore, during model development, diagnostics typically requires discrete intermittent samples of

data with known health states, while prognostics requires continuous samples of (run-to-failure) data, where the true health status is often partially known. According to Fernández-Delgado et al. [91], regression methods can generally be classified in several categories, such as the (generalized) linear regression, tree-based methods, such as regression trees, random forests, bagging and boosting, support vector regression, neural networks, and deep learning. A brief description of the methods, comparative analysis and implementation details, such as adjustable hyper-parameters can be found in the survey article by Fernández-Delgado et al. [91]. Linear regression is one of the simplest supervised machine learning technique, where a linear relationship between input and output data is assumed. Polynomial regressors are applicable in cases where the relationship between in- and output data is non-linear [92]. As the name suggests, regression trees are decision trees for regression problems. The non-leaf or internal nodes of a decision tree are logical expressions to be evaluated. The outcome of the evaluation results in further internal nodes or to a terminal (leaf) node. Decision trees typically begin at the top to bottom [92]. Bagging, random forest and boosting are tree-based ensemble methods which aims to improve the predictive performance via a combination of trees [92]. Bagging and random forests build on a random sample of data, where each tree is grown individually. An average of the individual regression trees is then evaluated for prediction. Boosting makes predictions based on a weighted sum of trees, which are grown sequentially [92]. Support vector regression is an extension of the support vector machine (SVM) originally developed by Vapnik et al. for classification tasks, in applications to regression problem [93], [94]. In general, it involves an optimization task of finding parameters of an hyper-plane that maximizes the margin between two classes [94]. (Artificial) neural networks consist of a number of processing elements called neurons. The neurons are interconnected, such that the nodes have an activation function and the edges are weighted. The weights are adjusted accordingly during training to obtain a relation between the input and target, while minimizing some cost function [95], [96], [97]. Deep learning can be seen as an advancement of machine learning, where information and consequently knowledge is discovered from raw data without a priori feature engineering. As opposed to a conventional shallow neural network that have a single level of representation for small data, a deep neural network possesses multiple levels of representations for large datasets [98], [99]. Typical deep learning methods employed for prognostics are autoencoders, long short-term memory (LSTM), convolutional neural network (CNN), and long short-term memory (RNN). An overview of the general applications of deep learning methods can be found in [99] and applications particularly for PHM can be found in the review article by Zhang et al. [100].

2.6 Diagnostics and prognostics evaluation methods

Irrespective of a diagnostics and prognostics task, model-based and history-based approach are applicable. Hence, a side-by-side comparison of model-based and history-based approach is presented in Table 2.2 to assist in selecting the suitable approach.

Table 2.2: Comparison of model-based and history-based approach, adapted from [64] and [67]

Criteria	Model-based approach	History-based approach
Physical understanding	Required	Beneficial
Physical degradation model	Required	Not required
Failure threshold	Often required	Often not required
Explainability and Transparency	High	Typically limited
Quantity of process history required	Low	High
Operating conditions	Often required	Beneficial
Modeling complexity	High	Medium
Generality and scope	Typically system-specific	System-independent
Accuracy	Dependent on system/process model	Dependent on data quality and quantity

In the following, commonly employed metrics in the PHM field are briefly elaborated on, to evaluate the performance of diagnostics and prognostics methods, with the aim of accessing the quality, that is, the similarity or closeness of the predicted quantities to the actual quantities.

Metrics for evaluating the performance of diagnostics methods

Depending on the employed approach, fault detection produces hard decisions, that is, a binary output of 0 (healthy) and 1 (faulty) or soft decisions, meaning values in the range $[0, 1]$, to allow some degree of healthiness and faultiness. Commonly employed metrics for diagnostics performance evaluation originate from machine learning classification. Since binary fault detection is considered here, focus is laid on binary-class classification metrics for hard decisions. Commonly employed classification metrics for hard decisions are based on the confusion matrix, also known as contingency table [101], while Brier score, cross-entropy and the area under the receiver operating characteristic curve (AUC) [102] are typically employed to evaluate soft classifier decisions. The confusion matrix and commonly derived evaluation metrics for a binary-class classification problem are exemplarily depicted in Figure 2.5. The confusion matrix is a square matrix where the ground truth, namely actual class instances are typically in the rows, while the predicted class

	Predicted positive	Predicted negative	
Actual positive	TP True positive	FN False negative	Sensitivity $TP/(TP+FN)$
Actual negative	FP False positive	TN True negative	Specificity $TN/(TN+FP)$
	Precision $TP/(TP+FP)$	Negative Predictive Value $TN/(TN+FN)$	Accuracy $(TP+TN)/(TP+TN+FP+FN)$

Figure 2.5: Confusion matrix and commonly derived evaluation metrics for binary-class classification

instances are typically represented in the columns. Correct predictions are located in the diagonal of the matrix, while incorrect predictions are in the off-diagonal. Depending on the location of the elements within the confusion matrix, the elements can be considered as true positive (TP), true negative (TN), false positive (FP), or false negative (FN). In diagnostics and specifically in fault detection, these quantities are commonly described as follows [103]:

- TP implies that a correctly *estimated faulty* unit is actually *faulty*,
- TN implies that a correctly *estimated healthy* unit is actually *healthy*,
- FP implies that an actually *healthy* unit is incorrectly *estimated faulty* and
- FN implies that an actually *faulty* unit is incorrectly *estimated healthy*.

FP typically results in costs-related consequences, because it results in false alarms, which in turn leads to increased economic costs due to unnecessary maintenance planning. Conversely, FN typically results in costs- and safety-related consequences, because an undetected fault have devastating human and environmental consequences, when it progresses into unplanned failure of safety-critical systems.

The following evaluation metrics are derived from the confusion matrix.

Sensitivity is also known as true positive rate (TPR), hit rate or recall [101], [104]. It is the ratio of the correctly predicted positive samples to the total number of positive samples. However, the metric does not consider negative samples, that is, neither TN nor FP [101], [104].

Specificity is also known as true negative rate (TNR) or inverse recall [101]. It is the ratio of the correctly predicted negative samples to the total number of negative samples. Similar, but conversely to sensitivity, the metric does not consider positive samples, that is, neither TP nor FN [101], [104].

On the one hand, as opposed to sensitivity where the sum of actual positives in the row is considered as the divisor, **precision**, also known as positive predictive value considers the sum of the predicted positives in the column as the divisor. On the other hand, as opposed to specificity where the divisor is the sum of actual negatives in the row, **negative predictive value** considers the sum of the predicted negatives.

Accuracy is the proportion of correctly classified samples. Although accuracy is easy to interpret and considers all observations, it is bias when the classes are imbalance. Class imbalance occurs when the instances of a class significantly dominate the instances of other class(es) [101], [105]. The class with larger percentage of instances is termed majority class, while the class with fewer instances is termed minority class [105]. In case of class imbalance, the **balanced accuracy** or **balance classification rate (BCR)** can be used. The BCR is the arithmetic mean of the sensitivity and specificity and is given as in Equation 2.1 [101]. Its value ranges from 0, as in worse to 1, as in perfect classification.

$$\text{BCR} = \frac{\text{TPR} + \text{TNR}}{2} = \frac{1}{2} \left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right) \quad (2.1)$$

A further metric that considers both the sensitivity and specificity is the **geometric mean (GM)** as defined in Equation 2.2, where a score of 1 indicates a very balanced performing model. The GM is also applicable in the case of class imbalance.

$$\text{GM} = \sqrt{\text{TPR} \cdot \text{TNR}} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FN}} \cdot \frac{\text{TN}}{\text{TN} + \text{FP}}} \quad (2.2)$$

BCR and GM are highlighted here because class imbalance is prevalent in most technical systems and specifically in safety-critical systems, where the minority class is the positive class or faulty state and the majority class is the negative class or healthy state. This imbalance is typically attributed to long operational periods with few or no recorded faults, leading to datasets dominated by normal behavior [106].

Prognostics performance evaluation metrics

In prognostics, the RUL is typically assigned a unit of time, such as in hours or cycles. However, it can also be assigned number of instances or measurements. The ground truth or actual RUL is typically not observable or directly measurable, but estimated, giving the current time t_k and the time at end-of-life (EoL) t_{EoL} . The EoL denotes the fault/failure threshold, which typically considers some safety margin [19]. The RUL at instance k is mathematically expressed as:

$$\text{RUL}_k = t_{\text{EoL}} - t_k, \quad (2.3)$$

where k typically begins at beginning-of-life (BoL), but not compulsorily. Generally, the performance of prognostics methods can be evaluated based on several criteria, such as algorithm prediction performance, computational performance, cost-benefit and ease of algorithm certification [107], [108]. Depending on the criteria, several metrics are conceivable, such as error metrics for algorithm prediction performance, the central processing unit (CPU) time for computational performance or the return on investment (ROI) for the cost-benefit [107], [108]. Error metrics are considered in the following to evaluate the performance of prognostics algorithms.

In a search for the best and widely employed metric to evaluate the performance of prognostics algorithms, a literature review of commonly employed error metrics, namely, the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE)

and the mean absolute percentage error (MAPE), was conducted with Google Scholar as the search engine. With slightly different results, a similar survey was conducted by Botchkarev [109] under consideration of four review articles between 1982 and 2007. In the last 30 years, the MSE remain the most commonly reported metric. According to Hyn-dman et al. [110], the popularity of the MSE and RMSE is attributed to their theoretical relevance, especially in statistics. Nonetheless, these metrics possess several limitations, which have led to their adaptation or the proposal of further metrics. For example, MSE and RMSE are more sensitive to outliers than MAE or MAPE [110]. The MAE is difficult to compare across datasets with different scales and the MAPE is undefined or infinite if the actual value is zero or close to zero [110]. Although the adaptation of the presented metrics, such as the geometric mean or median (relative) absolute error or the symmetric MAPE have solved some issues of the metrics, other limitations remain or emerge [111].

Specifically for prognostics, Saxena et al. [107], [108] proposed four sequential metrics to evaluate the performance of RUL predictions, namely prognostic horizon, $\alpha - \lambda$ performance, relative accuracy and convergence. The algorithm evaluation begins with the prognostic horizon, which evaluates whether an algorithm prediction is within a specified accuracy around the actual EoL. Thereafter, the $\alpha - \lambda$ performance qualitatively evaluates how well the algorithm performs within a predefined α -error margin at any time instant λ . The relative accuracy then quantifies the $\alpha - \lambda$ performance. On the premise that the three previous metrics are satisfied, convergence quantifies how fast the algorithm converges, as the name suggests. These sequential metrics also possess shortcomings, for example, the practical evaluation of the prognostic horizon might result in ambiguities, the $\alpha - \lambda$ performance is qualitative and hence does not facilitate in comparing different algorithms over all predictions [112]. The relative accuracy is dependent on a predefined α -error margin of the $\alpha - \lambda$ performance and lastly, it is difficult to compare the convergence across applications [112].

Thus, from the preceding and further literature survey on performance metrics [113], [114], it can be concluded that there is *no single best* metric to evaluate the performance of prognostics algorithms. In the following, the RMSE and the MAPE are considered in more detail to evaluate the prediction performance.

RMSE is one of the frequently employed metrics in the past 40 years, probably because it preserves the scale or unit of the predicted quantity and it is easily interpretable [110]. It is mathematically expressed as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^n (a\text{RUL}_k - p\text{RUL}_k)^2}, \quad (2.4)$$

where $a\text{RUL}_k$ and $p\text{RUL}_k$ are the actual and predicted RUL at instant k respectively, and n is the total number of predictions. Here $1 \leq k \leq n$ indicates the prediction interval, which necessarily does not always begin at BoL and end at EoL. To allow for comparison over several datasets, the RMSE is normalized by the mean of the actual quantities [113]:

$$\text{nRMSE} = \frac{\text{RMSE}}{\frac{1}{n} \sum_{k=1}^n a\text{RUL}_k}. \quad (2.5)$$

The output of normalized RMSE (nRMSE) is non-negative and values close to zero indicate perfect prediction.

MAPE is a commonly employed metric to evaluate the error between the predicted and actual quantities in companies [109], because it is easy to understand [111], [114] and its output is dimensionless or scale-independent, which allows comparison over different datasets [110]. It is mathematically expressed as:

$$\text{MAPE} = 100 \cdot \frac{1}{n} \sum_{k=1}^n \left| \frac{a\text{RUL}_k - p\text{RUL}_k}{a\text{RUL}_k} \right|. \quad (2.6)$$

From Equation 2.3, $a\text{RUL}_k = 0$ at instant $k = \text{EoL}$ results in an undefined output of the MAPE. Thus, the MAE is alternatively normalized by the mean of the actual quantities. This results in the weighted MAPE (wMAPE) [115] expressed as:

$$\text{wMAPE} = 100 \cdot \frac{\frac{1}{n} \sum_{k=1}^n |a\text{RUL}_k - p\text{RUL}_k|}{\frac{1}{n} \sum_{k=1}^n a\text{RUL}_k}. \quad (2.7)$$

The output of wMAPE is non-negative and the values close to zero indicate perfect prediction.

3 Accounting for time-varying operating conditions in PHM

To address the challenges posed by time-varying conditions in the development of diagnostics and prognostics of technical systems, their effects can be taken into account in the pre-data acquisition phase and in the post-data acquisition phase. The pre-data acquisition phase involves the technical understanding step of the extended PHM cycle, that is, before condition monitoring data is available. The post-data acquisition phase involves the data understanding, data preparation and modeling steps of the extended PHM cycle. In the following, individual methods within these phases are elaborated on, while critically surveying scholarly literature.

3.1 Consideration in the pre-data acquisition phase

Before data acquisition, domain experts can help to ascertain if it is feasible to acquire condition monitoring signals at predetermined constant and fixed operating conditions. Even if the system under consideration operates at time-varying conditions, the system might be specifically tunable to that fixed constant operating condition, where condition monitoring signals can be acquired at regular intervals. Although, this approach reduces the diagnostics and prognostics task to the case in which the operating condition is deterministic, it has two major limitations. Firstly, the approach is not generalizable, because the effects of varying operating conditions, such as sudden jumps or run-ups and run-downs that can lead to a significant change in the health condition or even sudden failure, are unconsidered. Secondly, the system's mode of operation might have to be interrupted for the data acquisition at the predetermined constant operating conditions.

Alternatively, domain experts can also help to uncover parameters of the system or process that are more sensitive to degradation than the effects of varying conditions. Even if the parameters of interest are not directly measurable, they might be derivable from directly measurable parameters, under supervision of domain experts. Dedicated instrumentation and measurement setup might be necessary for capturing degradation irrespective of varying conditions. For the diagnostics of bearings, as an example, the shock pulse method [116] or generally the shock-response spectrum analysis [117] at resonance frequencies of 32 kHz or higher either require the shock pulse meter or instrumentation capable of sampling at a frequency of at least two times the frequency of interest, specifically 64 kHz or higher. The shock pulse method is applicable for the diagnostics of bearings under varying speed and limitedly in varying load cases [118], [119]. However, the requirement of dedicated instrumentation and such relatively high sampling frequency, which presents a challenge in further analysis or data storage, are some of the demerits of such methods.

3.2 Consideration in the post-data acquisition phase

In the post-data phase, time-varying condition can be considered via specialized signal analysis techniques in the data preparation step, via condition-invariant feature engineering, feature scaling, feature standardization and feature segmentation. Physics-based models or empirical models, which consider the operating conditions can be employed in the modeling stage. Furthermore, the operating conditions can be considered as an additional input to a machine learning model.

3.2.1 Approach A: Signal analysis techniques for time-varying conditions

Although the involvement of domain experts is always beneficial in any phase of the diagnostics or prognostics process, the exploitation of expert knowledge at an early stage, such as in selecting the appropriate signal analysis technique is advantageous even in the subsequent stages. In the following sections, signal analysis techniques for rotating machinery are discussed in detail, because more research has been conducted in the field and given its significance to the present work.

After data acquisition, data exploration in the time-, frequency- or time-frequency domains can be conducted with a special focus on varying conditions. Where the shaft speed of a rotating machinery varies, order tracking can be employed for diagnostics. That is, with order tracking, faults are not only detected, but when the kinematics of the components under investigation are known, faulty components are correspondingly localized irrespective of variations in speed [120]. Order tracking is a frequency domain analysis method, where multiples of shaft rotating speed is the frequency base [121], [122]. Over the years, depending on if and how the shaft rotating speed is measured, order tracking methods have evolved from hardware order tracking, through computed order tracking to tacholess order tracking. As the name implies, in addition to acquiring the shaft speed and vibration signals, hardware order tracking utilizes a dedicated analog hardware to synchronize the vibration signal to be sampled at constant shaft increments [121], [122]. Limitations of this method are the hardware cost and its unreliability at rapidly varying shaft speeds with run-ups or run-downs [121], [122]. As opposed to utilizing a dedicated hardware, computed order tracking synchronizes the vibration signals with the shaft rotation speed through digital interpolation and re-sampling algorithms [121], [122]. In situations, where the shaft speed is not acquired or available, for example due to proprietary reasons, the tacholess order tracking method is applicable [123]. In tacholess order tracking, the shaft rotating speed is not directly measured, but derived from another measured signal, such as an acoustic or vibration signal [123]. Peeters et al. [124] broadly classifies existing methods for tacholess order tracking in methods based on signal demodulation and time-frequency representation. In their review article, a comparative study was provided for the speed estimation for three case studies, namely a wind turbine gearbox, an aircraft engine, and on a ship's generator. Although, the presented methods were generally capable of tracking the speed, a major challenge still remain when the shaft speeds rapidly varies [124]. Liu et al. [125] proposed an extension of the time-frequency representation method based on the short-time fractional Fourier transform to successfully diagnose a wind turbine gearbox subjected to time-varying speed. An overview of tacholess order tracking methods for the diagnostics of bearing faults can be found in the review article by Lu et al. [123]. Generally, shortcomings of the order tracking method

is that it is only applicable for rotating machinery, the kinematics of the components of interest are required for diagnostics and the effects of other varying conditions, such as the load is not directly considered. Furthermore, characteristic frequencies of interest can be masked as a consequence of smearing by neighboring frequencies, which introduces difficulties in fault identification and localization [124], [125].

Other commonly employed time-frequency domain analysis techniques are the wavelet and the Hilbert–Huang transform, which both accommodate the analysis of signals involving varying speed and load [126], [127]. Although the wavelet transform is one of the widely employed time-frequency domain techniques, there is no consensus on the optimal mother wavelet for analysis [126], [127]. Furthermore, the wavelet transform is computationally expensive and has a poor frequency resolution at high frequencies, thus impairing interpretation and fault identification [126], [127]. As opposed to the wavelet transform, the empirical mode decomposition of the Hilbert–Huang transform is not as computationally expensive, therefore it is better suited for signals with high sampling rates [127]. Nevertheless, a major limitation of the empirical mode decomposition is that it introduces aliasing, which impedes fault identification and interpretation [127]. Although several extensions or adaptations of the wavelet and Hilbert–Huang transform have been proposed, they have either increased complexity or decreased interpretability [127].

3.2.2 Approach B: Condition-invariant feature engineering

With a special focus on time-varying operating conditions, condition-invariant feature engineering involves either the manual or automatic extraction and subsequent selection of features that are irrespective of the effects of time-varying conditions, but sensitive to degradation effects. Such condition-invariant features or feature subset obviates the need for acquiring the operating operations, thus reducing the efforts and costs of such sensors. To this end, there have been several attempts in research to find possible condition-invariant features, ranging from time-domain features to deep features. For example, Hou et al. [128], [129] proposed several features derived from quasi-arithmetic means and adapted sparsity measures based on adaptive weighted signal preprocessing technique as alternatives to statistical features, such as the root mean square (RMS), which are typically susceptible to the influence of operating conditions. Under consideration of experimental accelerated bearing degradation datasets, the proposed features were proven to be better health indicators than statistical features. However, their effectiveness under the influence of time-varying operating conditions was unconsidered. Singh et al. [130] proposed a confidence value indicator derived from the Jensen–Rényi divergence of an ensemble EMD, where the Jensen–Rényi divergence is calculated based on the relative probability distribution of the energies of the intrinsic mode functions. However, the proposed indicator was only evaluated for cases where the load or speed of a rotating machinery vary up to 10%. Furthermore, in addition to the shortcomings of the EMD, the α parameter of the Jensen–Rényi divergence requires expert definition. Schmidt et al. [131], [132] utilized the time-frequency methods of the instantaneous power spectrum and the cyclic modulation spectrum to construct features, such as the maximum value of the optimal frequency bands of the kurtogram, infogram, sparsogram, ICS2gram, log-cycligram and IFBI $_{\alpha}$ gram. Their proposed methodology was evaluated based on experimental data of gearboxes to detect local and distributed gear faults. However, their methodology requires the prior knowledge of the rotating components. Wang et al. [133] employed a transfer learning method based on joint dictionary learning matrix factorization to learn

condition-invariant features from scalogram images derived from the continuous wavelet transform. However, the proposed transfer learning method was evaluated on the pronostia dataset [134], where the operating condition is constant within each experiment and only change between different run-to-failure experiments. Rombach et al. [135] employed the triplet loss of the contrastive learning method to extract deep features. Although the learned feature representation successfully allowed the classification of known faults and the clustering of unknown faults, the proposed method was evaluated on the case western reserve university (CWRU) bearing dataset, where the operating condition only change between different run-to-failure experiments. Michau and Fink [136] learned deep features from the encoder component of an adversarial deep learning method. The learned features in one operating condition domain were successfully transferred to another operating condition domain employed to detect faults in an unsupervised manner for several case studies. However, in addition to the sufficient amount of data required to train deep models, the operating condition for training and testing was collected under the same degradation level, which requires that the data covers all possible situations.

On the whole, several attempts have been made to find possible condition-invariant features. However, most features derived from time-frequency representation typically require expert knowledge, while deep features derived from autoencoders typically require sufficient amount of data. Furthermore, most of the attempts have either increased complexity or decreased interpretability [127]. Since condition monitoring signals are typically time series data, time series analysis toolboxes can be considered in a further attempt to find condition-invariant features. The utilization of such time series analysis toolboxes is expert-independent and does not depend on machine learning algorithms. In the following, the major findings of an article [137], which investigates time series analysis toolboxes for the condition-invariant feature engineering are briefly described.

The EMG feature extraction toolbox and the HCTSA package [36], [37] were investigated alongside statistical time-domain analysis methods to extract features from vibration data. The EMG feature extraction toolbox is dedicated for the extraction of features from electromyograms [38], [39] and thus has its roots in the medical field. An electromyogram is a time series waveform measured for the diagnostics of muscles. Therefore, since most condition monitoring signals, such as vibration signal, are also time series waveform, the EMG feature extraction toolbox was considered to extract 17 features. Furthermore, some researchers [138], [139] successfully employed EMG features for the fault diagnostics of rotating machinery components, namely gearbox and bearing. Although various fault types could be successfully diagnosed under varying operating conditions, the operating conditions were kept constant per experiment in their study. The HCTSA is a comprehensive toolbox not only for the extraction of features from time series, but also for the visualization and analysis of time series [36], [37]. As opposed to the EMG feature extraction toolbox that comprises 40 features [140], the HCTSA package comprises 1000 basic operations, which can result to more than 7700 characteristic features [36], [37]. Furthermore, in addition to features from the medical field, the HCTSA package is a collection of features from interdisciplinary fields such as economics and physics. Although the HCTSA package has found numerous applications in the medical field [37], such as for classifying heartbeat [141] and stress level [142], it has found limited application for diagnostics or prognostics task, while considering vibration data [37]. To identify significant condition-invariant feature subset from the interdisciplinary extracted features, the chi-square test and monotonicity were employed for diagnostics and prognostics task, respectively, because of their independence on machine-learning techniques [42]. The

Chi-square test is a statistical test of the independence of two statistical variables, where features that are dependent of the target class are more relevant [138]. Here, the target class consists of an healthy and an unhealthy class for fault detection problem and consists of the faulty classes for fault identification problem. Monotonicity was coined by Coble and Hines [143] and defined based on the derivatives of time series to imply an increasing or decreasing trend. As opposed to their definition, the Spearman's rank correlation coefficient, which evaluates the general monotonic-relationship is considered, because it was successfully employed in prognostics [144]. The described methods were successfully evaluated based on two case studies [137]. One case study considered the effectiveness of the proposed methods for mapping intermittent samples of vibration data collected under different time-varying shaft speed profiles and discrete health states to discrete faults of bearing components. The other case study considered the effectiveness of the proposed methods for mapping continuous samples of vibration data acquired from a real-world wind turbine subjected to time-varying wind speed to a continuous health indicator.

3.2.3 Approach C: Condition-centered data scaling or feature scaling

In a narrow sense, condition-centered data scaling or feature scaling is the ratio of data or a feature derived from a condition monitoring signal to data or a feature derived from an operating condition at a given time instant, with the aim of homogenizing and factoring out the effects of the operating condition. In a broader sense, it also involves utilizing baseline measurements, often healthy, acquired at different, but specific operating conditions of interest to compensate for the influence of the operating condition, for example through arithmetic operations. In the following, a literature survey from the PHM and SHM field is presented to identify possible limitations of current research.

Under the assumption that the amplitude or frequency modulation resulting from a fault is separable from that of load-induced amplitude or frequency modulation, Stander and Heyns [145] proposed a load demodulation normalization technique to mitigate the effects of varying loads on vibration data. In summary, the vibration data is scaled by the filtered modulation signal representing the load effects. The proposed normalization technique was successfully employed to diagnose different levels of faults on gears within an experimental single-stage gearbox subjected to different time-varying load profiles. However, shortcomings of the proposed normalization technique is that it requires an optimization procedure to determine the optimal filter frequency bandwidth for demodulation. Furthermore, it attenuates the fault component of the signal and hence it can even complicate effective diagnostics or prognostics [146]. To avert the attenuation property of the load demodulation normalization technique resulting from the finite impulse response filter, Schmidt and Heyns [146] proposed the utilization of moving median windowing procedure. In the SHM of reinforced concrete structures, Chrisp et al. [147] considered the changes in the electrical conductivity to indicate the deterioration of the concretes. They considered the normalization of the concrete conductivity measurements through a so-called temperature correction formula to a predetermined reference temperature, typically between 20 °C and 25 °C. Jha and Dong [148] also employed the same temperature correction formula to normalize resistance measurements for the temperature-compensation in monitoring the degradation of power electronics, such as metal-oxide-semiconductor field-effect transistors (MOSFETs). However, Chrisp et al.

[147] found out that a major limitation of the temperature correction formula is that it is typically valid within $\pm 5^\circ\text{C}$ around the reference temperature. Thus, they [147] proposed the Arrhenius relationship to normalize the concrete conductivity. However, this introduces further parameters, such as the activation energy for the conduction process, which is concrete-dependent and requires experimental determination [147]. Rigamonti et al. [149] measured the capacitor equivalent series resistance to monitor the degradation of aluminum electrolytic capacitors installed in full electric vehicles. To compensate for the effects of varying operating temperatures, Rigamonti et al. [149] considered the ratio of the measured capacitor equivalent series resistance at an arbitrary temperature to the initial capacitor equivalent series resistance at the same temperature. Their approach is similar to the widely employed optimal baseline selection methodology for temperature compensation from the SHM community [150], which involves establishing and comparing baseline measurements to subsequent measurements. Furthermore, the approach by Rigamonti et al. [149] can also be considered a realization of the ISO 17359 [26] guideline, wherein it is proposed to acquire baseline at the distinct operating conditions and compare subsequent measurements to such baselines. However, as previously stated, such baseline-dependent techniques are only directly applicable in situations where the operating conditions are distinct and where there is a clear transition between changes in the operating condition. Furthermore, such baseline databases might be impractical for cases where the future operating condition profile is not a priori known.

A major limitation identified from the presented literature survey is that condition-centered data or feature scaling is typically employed to consider the effects of a single operating condition.

3.2.4 Approach D: Condition-specific feature standardization and feature segmentation

Most machine learning techniques, such as neural networks, perform better when the input data is normalized or standardized [151]. Thus, standardization is typically employed to normalize (input) features as a pre-processing step in the application of machine learning techniques. In the light of varying operating conditions, features are not globally standardized for the purpose of applying a specific machine learning, but standardized per distinct operating condition to compensate the effects of varying operating conditions.

For the first data challenge of the PHM society [152], national aeronautics and space administration (NASA) provided turbofan engine degradation datasets subjected to time-varying operating conditions, which were simulated with the commercial modular aero-propulsion system simulation (C-MAPSS) software. The dataset comprises 218 engine units, each with 21 sensor measurements and corresponding three operational settings. Via different visualization techniques, the winning team of the IEEE GOLD category [153] identified that the three operational settings resulted in six distinct regimes or modes of operation. Thus, to successfully tackle the task of predicting the RUL of the turbofan engines, each feature value X_k was standardized per operating mode [153], according to the following equation.

$$\tilde{X}_k = \frac{X_k^{o,f} - \mu^{o,f}}{\sigma^{o,f}}, \forall o, f \quad (3.1)$$

where $\tilde{X}_k$ is the standardized feature at instance $k \in \{1 \dots n\}$, $o \in \{1 \dots m\}$ is the index of operating modes, $f \in \{1 \dots p\}$ is the feature index, μ denotes the mean, σ denotes the standard deviation, n , m , and p denote the total number of instances, number of distinct operating modes and number of features, respectively. Alternatively, the overall winning team [154] segmented selected features into the distinct operating modes, also identified via data visualization. Thereafter, each feature segment was employed to fit an individual regression model, yielding a total of six models for the total number of distinct operating modes. Further approaches have been developed around the previously described approaches by Peel [153] and Wang et al. [154], also based on the C-MAPSS dataset. For example, Xu et al. [155] employed k-means clustering algorithm to identify the operating modes of the C-MAPSS dataset. Subsequently each feature value was standardized per operating mode, according to equation 3.1 and the resulting standardized features were min-max normalized to the $[0, 1]$ interval. Finally, the standardized and min-max normalized features were recombined with the operating condition information to be fed as inputs to a machine learning model to account for possible future operating conditions. Javanmardi et al. [156] also clustered the operating modes of the C-MAPSS dataset with the k-means clustering algorithm, but min-max normalized the sensor measurements within each operating mode to the $[-1, 1]$ interval without prior standardization. Since the number of operating conditions or resulting operating modes are typically not known in advance, as required by the k-means clustering algorithm, Baptista et al. [157] utilized a self-organizing map to group the distinct operating modes of the C-MAPSS dataset into respective clusters without predefining the number of expected clusters. Furthermore, Baptista et al. [157] proposed a so-called normalizing multi-layer perceptron neural network to normalize the sensor data. Inputs to the network were the non-normalized sensor data and the factorized operating mode cluster. Furthermore, there is no target fed to the network, but the network works in an unsupervised manner to minimize the square of predictions, while adjusting the network weights [157]. Zhang et al. [158] employed density peaks clustering algorithm and hierarchical density-based spatial clustering of applications with noise algorithm to identify the operating modes of the C-MAPSS dataset without predefining the number of clusters. Zhang et al. [158] proposed the multi-operational condition-based normalization method, which is actually based on equation 3.1, to standardize the features. Finally, as performed by Xu et al. [155], the standardized features were recombined with features from the operating condition. To derive an health indicator for rolling element bearings subjected to a stepped cyclic speed and loading profiles, Kim et al. [159] initially clustered the operating conditions with the k-means clustering algorithm into two modes of operation. Thereafter, statistical features derived from vibration data were partitioned and a subset of the derived features were selected per cluster according to the monotonicity score to construct the health indicator. To account for time-varying conditions of rolling element bearings subjected to time-varying speed and loading conditions, Sim et al. [11] proposed a derivative-based approach to partition the conditions into finite number of distinct operating conditions. The distinct operating condition regions were then employed to manually partition raw vibration data before extracting features per segment.

The main assumption laid for the condition-specific feature standardization and feature segmentation by the surveyed literature is that the operating conditions are distinct and finite or separable in finite number of distinct operating conditions. Furthermore, it is typically assumed that the operation condition occurred with identical probability at each cycle [157]. Both assumptions are not always met, for example in cases where the

time-varying operating condition is random or transient.

3.2.5 Approach E: Embedding in the modeling phase

Embedding in the modeling phase often involves the incorporation of the operating conditions within model-based models or as input variables in machine-learning models to compensate for the effects of time-varying operating conditions. In the following, literature considering single and multiple operating conditions and considering different modeling approaches are surveyed.

Roy et al. [160] developed a physics-based model to compensate for the effects of varying temperature in SHM applications. Through numerical simulations based on spectral element method, the influence of varying temperature on the material properties of piezoelectric transducers was investigated. Based on the identified relationship, the piezoelectric sensor signal could be reconstructed, which filters out the influence of the varying temperature during fault diagnostics. Xu et al. [8] considered the Arrhenius and Wiener process model to consider the influence of time-varying temperature conditions on the degradation of lithium-ion batteries. The model parameters were estimated based on the maximum likelihood estimation method in combination with genetic algorithm. Subsequently, the RUL was predicted based on the Bayesian framework. The proposed methodology was validated with a lithium-ion battery in a temperature chamber subjected to temperature changes between 25 °C and 55 °C at 10 °C step during charge-discharge cycles.

Under the premise that the degradation rate and the variation of the degradation process are jointly dependent on the operating conditions, Wang et al. [161] utilized two jointly dependent functions, namely a state function, described by a Wiener process to model the degradation process and an observation function that describes the measured degradation signal. The model parameters were estimated via a combination of the maximum likelihood and least squares estimation, while the RUL was predicted by a Bayesian framework. The proposed methodology was successfully evaluated on rolling-element bearings, which were subjected to a constant load amplitude and three-level stepped speed conditions. Li et al. [162] proposed a so-called two-factor state-space model-based method to predict the RUL under consideration of time-varying operating conditions. It also involves the Wiener process as the degradation process model and a measurement function, where the model parameters were estimated via the maximum likelihood estimation method. The proposed method is built on the assumption that changes in operating conditions result in signal jumps, which are considered in the measurement function via a signal transformation algorithm. Li et al. [162] validated their method via accelerated degradation tests of roller thrust bearings subjected to constant load and varying speeds of 2200 rpm and 2600 rpm. Yucesan et al. [163] considered a physics-informed neural network based on the bearing basic rating life and a recurrent neural network to model the degradation of the main bearing in a wind turbine. They adapted the bearing basic rating life via the Palmgren-Miner's rule to characterize incremental damage of a bearing subjected to varying load and speed at each cycle. Furthermore, the viscosity and contamination level of the bearing grease were also considered in their model via a multi-layer perceptron neural network. Their proposed methodology was successfully evaluated on simulated wind turbines with varying speed profiles. Geurts et al. [164] also considered the basic life rating of a bearing, which incorporates the bearing load and speed, in their hybrid model. During the healthy stage of a bearing, the basic life rating is utilized to estimate the RUL, while a Bayesian-based method is employed to predict the RUL, when a fault is

detected. Their methodology was evaluated on accelerated tests of ball bearings subjected to varying speed of 1000 rpm to 2000 rpm in steps of 100 rpm.

Kayedpour et al. [165] proposed an hybrid health monitoring framework for wind turbines subjected to time-varying speed conditions. The parameters of the wind turbine, such as the pitch-generator torque, the electrical power and wind-dependent system response were physically modeled and integrated into a deep learning surrogate model to learn the relationship between the system degradation and the effects of operating condition. Shi et al. [166] proposed a physics-informed LSTM model that integrates a physics-based calender and cycle aging model within an LSTM network. The physics-based portion accounts for the degradation under consideration of the operating condition, while the LSTM considers the time-dependent effect of the health state on the degradation. The proposed physics-informed LSTM model was evaluated on batteries subjected to different temperature and load profiles. Long et al. [9] proposed a stochastic hybrid model, which consists of a Wiener model and a continuous-time Markov chain model, to model the degradation process under time-varying operating conditions. Furthermore, they employed the particle filter to predict the RUL. Long et al. [9] evaluated their methodology on bearings and motorized spindles of CNC machine tools subjected to varying speed and load under three operational settings.

The shortcomings of embedding time-varying operating conditions in the modeling phase are inherited from the underlying methodology, that is, it depends on the employed method. Furthermore, the complexity of the underlying methodology might even be heightened, such as for machine learning models, where the operating conditions increase the number of inputs. The increase in model inputs might diminish accuracy when the inputs are of low quality, it leads to an increase in the number of required training data, it impedes interpretability and explainability, it leads to an increase in computational complexity and subsequently to an increase in required resources [167], [168].

3.3 Framework for fault detection and prognostics under time-varying operating conditions

From the preceding section, most of the existing literature on diagnostics and prognostics under time-varying operating conditions consider a specific methodology without a general guideline or framework for considering the effects of time-varying operating conditions. Furthermore, they either consider the variation of a single variable, such as the shaft rotating speed during accelerated lifetime tests of rotating machinery components, such as bearings or gears or they consider a very limited amount of operating modes, as in the C-MAPSS data with six operating modes. To generalize the consideration of time-varying operating conditions and address real-world scenarios, a general framework for considering the effects of time-varying operating conditions is presented in this section, while a case study considers the random variation of multiple operating conditions in a subsequent section, to address the research gap. In the following, condition-related data is used interchangeably with condition monitoring data, which involves data pertaining to the health state of the monitored system.

Figure 3.1 presents a general procedure for data acquisition and data preparation of operating condition data and condition monitoring data. The prepared operating condition and condition monitoring data then serve as inputs to the framework presented in Figure

3.2, which generally considers time-varying operating conditions in diagnostics and prognostics. In Figure 3.2, the model represents an arbitrary fault detector or RUL estimator, while the target represents the respective model output, as in discrete health state or continuous RUL value. The presented framework should serve as a guide for novice and practitioners in the development and application of diagnostics and prognostics algorithms for technical systems, with a special focus on considering the effects of time-varying operating conditions. The flowchart builds on the availability of condition monitoring data and where applicable, synchronously acquired operating condition data. In approach A, specialized signal analysis techniques, such as described in Section 3.2.1 can be employed to uncover the effects of the the operating conditions under consideration of condition monitoring data and optionally the operating condition data, before subsequent feature engineering and the modeling step. In approach B, if the operating condition data is unknown or not acquirable, due to reasons associated with accessibility or costs, then the applicability of features that are invariant of the operating condition, as described in Section 3.2.2, can be investigated, before applying any model-based or history-based modeling methods. If the operating condition data is available, then features derived from the condition monitoring data can be scaled, standardized or segmented by the operating condition, as described in Sections 3.2.3 and 3.2.4, for approaches C and D, respectively. During feature scaling, standardization or segmentation, it might be necessary to identify a single operating condition that dominantly influences the acquired condition monitoring signals. Alternatively to approaches C and D, features derived from the operating conditions can be embedded alongside the condition monitoring features in the modeling phase, as described in Section 3.2.5 for approach E. Although not represented in Figure 3.2, the proposed approaches can be combined. For example, features derived from specialized signal analysis techniques, such as the order tracking method or features invariant to a certain condition, such as the speed, can be combined with other operating condition monitoring data, such as the load. They can then serve as inputs to machine-learning techniques to account for the unobserved or unaccounted operating condition.

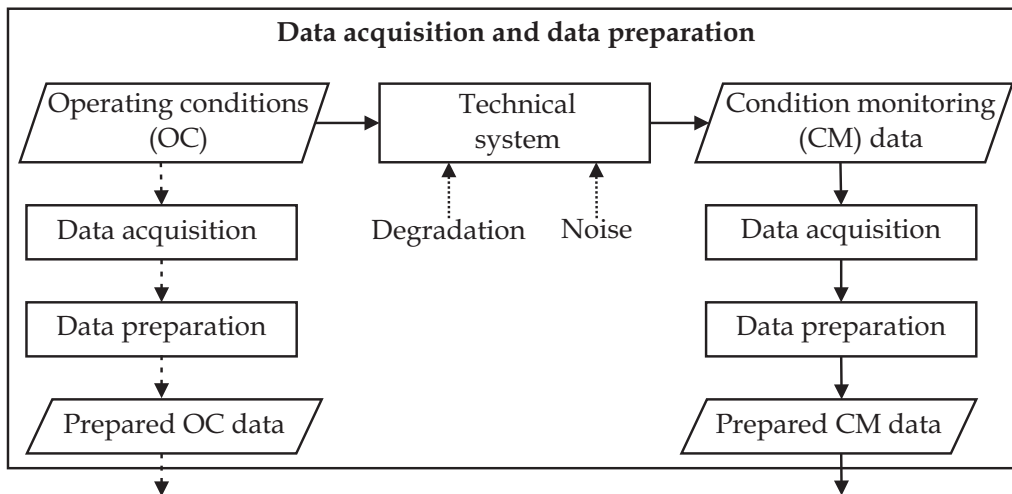


Figure 3.1: General procedure for acquiring and preparing operating condition and condition monitoring data

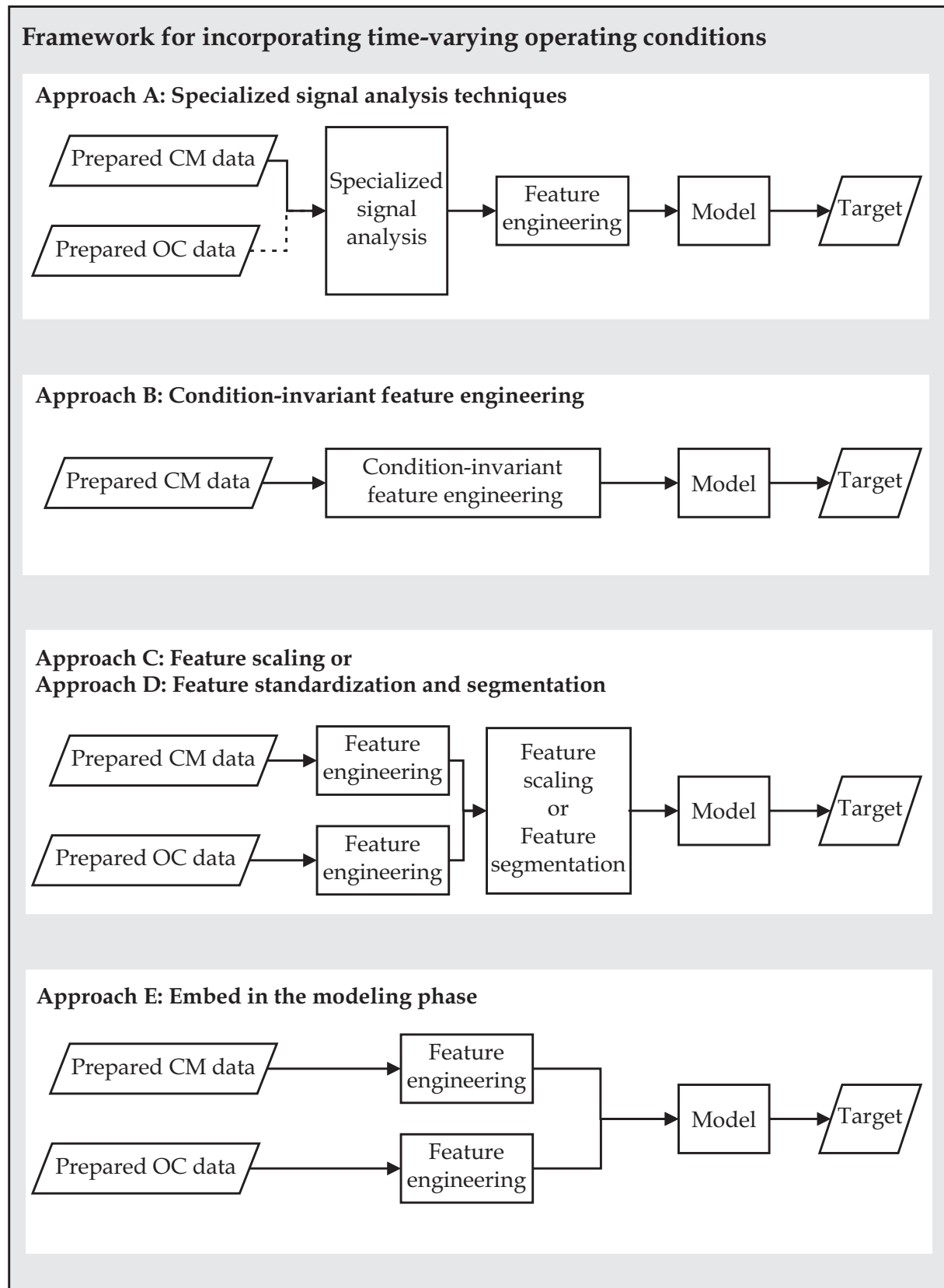


Figure 3.2: Flowchart for considering time-varying operating conditions in diagnostics and prognostics, adapted from [169]

4 Fault detection and prognostics of piezoelectric bending actuators

Piezoelectric materials exhibit both direct and converse piezoelectric effects. That is, they generate an electrical charge or voltage under pressure, when considering the direct piezoelectric effect, while they mechanically deform or produce strain under an applied electric field, when considering the converse piezoelectric effect. Thus, piezoelectric materials can be employed as sensors, actuators or transducers [170]. Furthermore, due to the absence of mechanical wear parts and other characteristics such as miniaturization capability, piezoelectric materials are utilized in numerous mechatronic applications, such as in medical dosing systems, energy harvesting, textile knitting machines, automotive fuel injectors, wire bonding, welding, and SHM applications [171], [172], [173], [174], [175]. However, due to the brittleness and typically low fracture toughness of the piezoelectric ceramics, they are susceptible to fault and subsequently to failure. Furthermore, defects, such as microcracks, inclusions and pores can arise from geometrical imperfections or material inhomogeneities, which can further propagate into failure under combined electrical, mechanical and/or thermal load [176], [177]. Accordingly, diagnostics and prognostics serve to mitigate unplanned failure and to enhance the long-term availability and reliability of piezoelectric materials utilized in sensing, actuation, and transduction applications [175].

4.1 Technical understanding

In the following, focus is primarily laid on piezoelectric bending actuators, such as shown in Figure 4.1. They can be found in Braille pads, medical dosing systems, textile knitting machines, micro-robots, positioning and assembling machines [171], [172], [173], [174], [175], [178]. Piezoelectric bending actuators typically consist of one or more piezoelectric layers with or without a bonded passive layer, which undergo bending motion depending on the applied electric field with respect to the polarization direction [179], [180]. Piezoelectric bending actuators are specifically termed uni-/monomorphs with one piezo-

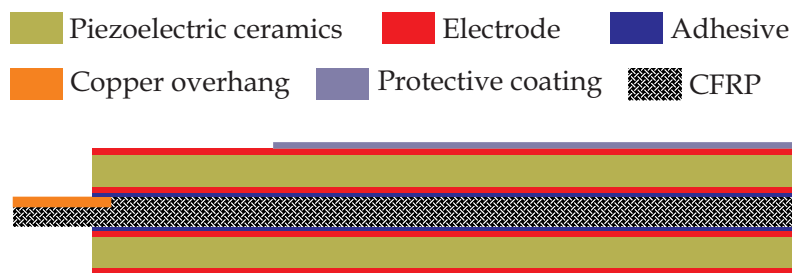


Figure 4.1: Side view showing the layers of the piezoelectric bending actuator adapted from [175], [178].

electric layer and a passive layer, bimorphs with two piezoelectric layers, trimorphs with two piezoelectric layers and a passive layer and multi-/polymorphs with more than two piezoelectric and passive layers [179]. The trimorph with two piezoelectric ceramics, four electrodes, a conductive passive middle layer based on carbon fiber reinforced polymer (CFRP), adhesive and a copper overhang to electrically contact the inner electrodes, as shown in Figure 4.1, is considered here. Based on the supporting paper [175], lifetime tests published in [181] and on reviewed literature [182], [183], [184], [185], [186], [187], [188], [189], [190], the FMEA of the trimorph layers are listed in Table 4.1.

Because no significant failure mode was recorded in the literature, the copper overhang is not considered here. As part of the risk analysis [186], failure modes associated with the piezoelectric ceramics, electrodes, passive middle layer, and adhesive are assessed in terms of their importance and the likelihood of their occurrence. All failure modes are equally important, since they negatively impact the actuator's performance and its overall functionality. However, because leakage current in the piezoelectric ceramics can potentially compromise safety, for example, by additionally causing damage to the control electronics due to an electrical short circuit, it is deemed more important than the other failure modes [175]. Crack development in the piezoelectric ceramics or electrodes is very probable, due to the brittleness of the piezoelectric ceramics and as described by Ruschmeyer [184]. Although, electrode burns are very probable, according to Denzler's study [183], they are not very probable here, because of the significantly lower voltage amplitudes, that is, approximately 40 V alternating current (AC) in the application example described here, as compared to the 200 V applied in Denzler's experiments. Depolarization is considered unlikely, provided the actuator is operated correctly. Cracks in the passive middle layer are also unlikely, due to the relatively higher toughness of the CFRP than the piezoelectric ceramics and the relatively low voltage amplitudes and resulting deflection. Adhesive delamination is not well-documented in existing literature, hence it is difficult to assess its probability. Soldering failure and leakage current are not anticipated in the controlled laboratory environment. In conclusion, cracks in the piezoelectric ceramics and electrode layers are deemed the most relevant failure modes for the application example [175]. Thus, focus is laid on diagnosing these failure modes.

Table 4.1: FMEA of the piezoelectric bending actuator [175]

Failure mode	Failure effect	Failure cause
Piezoelectric ceramics		
Crack	Reduced stiffness, increased resistance, consequent crack in the electrode	Mechanical load
Depolarization	Decreased piezoelectric characteristics, flatter resonance peaks, resonance and anti-resonance frequency coincide, loss of functionality	Operation with an electrical voltage opposite to the actuator's polarization, mechanical load opposite to the actuator's polarization, elevated temperature
Leakage current	Overload of electronic components in the measurement chain, such as the piezoelectric amplifier, loss of functionality	Elevated unipolar electric field strength, electromigration, contamination, humidity
Electrode		
Crack	Increased contact resistance, increased thermal load, reduced capacitance, decreased active area	Crack of the piezoelectric ceramics, mechanical load
Burn	Reduced capacitance, decrease in the active area	Thermal load
Passive middle layer		
Crack	Altered damping and stiffness characteristics	Mechanical load
Copper overhang		
Soldering failure	Partial or complete loss of electrical connection and loss of functionality	Improper soldering, mechanical vibrations, thermal load
Adhesive		
Delamination	Decrease in bending movement, loss of functionality	Mechanical load

Condition monitoring parameters for the piezoelectric bending actuator

The input and output parameters of the piezoelectric actuator are considered to select appropriate measurement techniques for the failure modes deemed most relevant. Owing to the piezoelectric effect, the piezoelectric bending actuator is an electromechanical system. Hence, it can be modeled via a four-pole model, which consists of an electrical and mechanical port [191], [192]. The electrical port encompasses an oscillating voltage $u(t)$ and current $i(t)$, while the mechanical port encompasses velocity $v(t)$ and force $F(t)$, as illustrated in Figure 4.2(a). The corresponding lumped-parameter model of a piezoelectric actuator actuating near its fundamental resonance frequency is shown in Figure 4.2(b). The electrical capacitance and resistance are denoted by C_e and R_e , respectively. Furthermore, the effective vibrating mass of the actuator is denoted by m_m , the internal mechanical damping is denoted by d_m , the mechanical stiffness is denoted by c_m and the electromechanical coupling factor is denoted by κ [175].

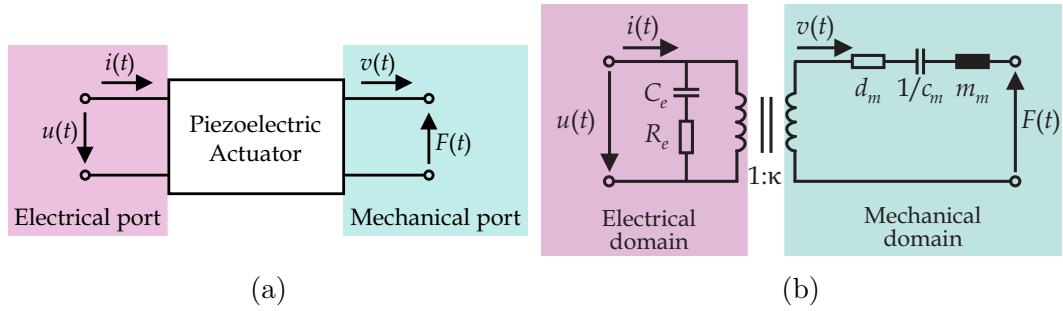


Figure 4.2: (a) Four-pole model adapted from [192] and (b) corresponding lumped-parameter model of a piezoelectric actuator adapted from [175], [191] .

The coupling matrix, which describes the relationship between the inputs and the outputs of a piezoelectric actuator is given in Equation 4.1, such that the input vector consists of the complex amplitudes of the voltage $\check{\underline{u}}$ and force $\check{\underline{F}}$ and the output vector consists of the complex amplitudes of the current $\check{\underline{i}}$ and the velocity $\check{\underline{v}}$. Furthermore, the actuating angular frequency is denoted by ω and the complex amplitude of the harmonic signals is denoted by the check accent mark [175], [191], [193].

$$\begin{bmatrix} \check{\underline{i}} \\ \check{\underline{v}} \end{bmatrix} = \begin{bmatrix} \frac{\kappa^2}{j \cdot \omega \cdot m_m + d_m + \frac{c_m}{j \cdot \omega}} + \frac{1}{R_e + \frac{1}{j \cdot \omega \cdot C_e}} & \frac{\kappa}{j \cdot \omega \cdot m_m + d_m + \frac{c_m}{j \cdot \omega}} \\ \frac{\kappa}{j \cdot \omega \cdot m_m + d_m + \frac{c_m}{j \cdot \omega}} & \frac{1}{j \cdot \omega \cdot m_m + d_m + \frac{c_m}{j \cdot \omega}} \end{bmatrix} \cdot \begin{bmatrix} \check{\underline{u}} \\ \check{\underline{F}} \end{bmatrix}. \quad (4.1)$$

For an unloaded piezoelectric actuator, that is, when $\check{\underline{F}} = 0$, the current is directly proportional to the applied voltage. Therefore, in a healthy actuator, increasing the driving voltage amplitude leads to a higher current amplitude and a lower resonance frequency, while decreasing the voltage amplitude results in a lower current amplitude and a higher resonance frequency. Therefore, any deviation in the current that is not proportionally reflected in the voltage indicates a change in the actuator's internal parameters, potentially due to a crack [175].

According to the results of the FMEA presented in Table 4.1, the presence of a crack in the electrodes leads to a decrease in capacitance, while cracks in both the electrodes and the piezoelectric ceramics cause an increase in resistance. Additionally, a crack in the piezoelectric ceramics reduces the mechanical stiffness. Capacitance and resistance

can be derived from the electrical admittance Y_e or impedance measurements Z_e , whereas mechanical stiffness can be determined from mechanical admittance Y_m or impedance Z_m [193], [194]. Electrical impedance is a complex quantity defined as the ratio of the applied voltage to the resulting current, while the mechanical admittance is a complex quantity defined as the ratio of the applied force (or applied voltage) to the resulting velocity [172], [194]. Admittance is the reciprocal of the impedance and is typically measured using an impedance analyzer across a range of frequencies.

Avvari et al. [195] intermittently measured the impedance of a piezoelectric energy harvester to monitor crack propagation. Their experimental results demonstrated that the resonance frequency decreased as crack progressed, due to stiffness reduction. Hensel et al. [173] also confirm that the presence of a crack in a power transducer leads to a reduction in stiffness, which in turn causes a decrease in both the resonance and anti-resonance frequencies of the measured electrical admittance. Kimotho et al. [174] also intermittently measured the electrical admittance of piezoelectric bending actuators during lifetime tests, to monitor degradation. Their findings indicated that as degradation progressed, the maximum absolute value of electrical admittance decreased, while the equivalent resistance increased. They also pointed out a key limitation of admittance measurements, which is that such measurements generally require interrupting the actuator's operation.

General experimental procedure

The test rig developed by Kimotho et al. [174] and Bender et al. [181] is adopted here to conduct run-to-failure experiments of piezoelectric bending actuators, as exemplarily shown in Figure 4.3(a). The piezoelectric bending actuators utilized in the experiments are coated parallel trimorphs with a conductive passive CFRP middle layer, such that the two piezoelectric layers have their polarization directions aligned. The bending actuators are manufactured by Johnson Matthey (Type1 - 427.0085.11F) and have a length of 50 mm, a width of 7.2 mm and a thickness of 0.81 mm. Further characteristics of the bending actuator can be found in the Appendix in Table 1. During the experiments, the bending actuators are clamped between two aluminum bars to ensure the electrodes on both piezoelectric ceramics are on the same potential. When clamped as in a cantilever, the effective length is approximately 38 mm, while the distance between the clamp and the protective coating on the piezoelectric layers is approximately 1 mm, as depicted in Figure 4.3(a) and Figure 4.3(b), respectively. In the following, the terminology cantilever piezoelectric bending actuator or trimorph is referred to as bending actuator, for brevity.

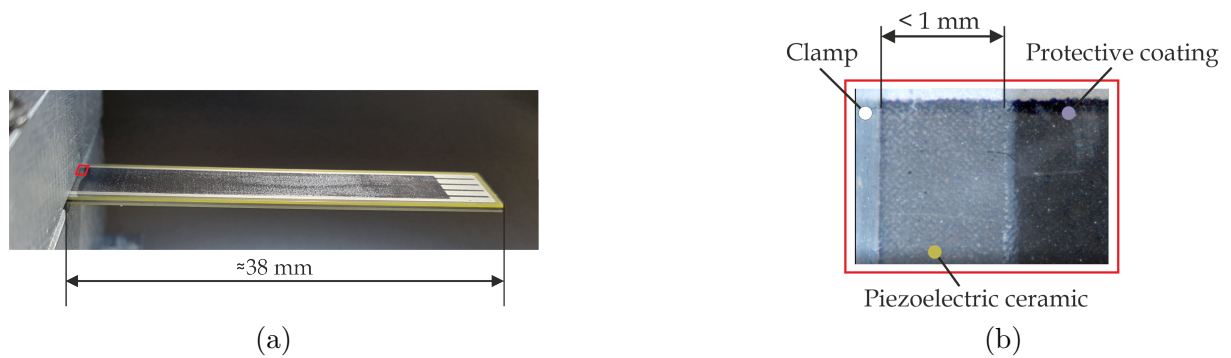


Figure 4.3: Piezoelectric bending actuator from Johnson Matthey (a) clamped between two aluminum bars and (b) zoomed portion of the top view showing probable region of crack occurrence [175].

The main purpose of the experiment is to accelerate the degradation of the bending actuators under time-varying operating conditions, while acquiring condition monitoring signals. In reference to the FMEA, the most relevant failure modes are cracks in the piezoelectric ceramics and electrode layers. Furthermore, since a crack in the piezoelectric layer accordingly implies a crack in the electrode, the goal of the experiment is to accelerate the development of a crack in at least a piezoelectric layer of the bending actuator. In most applications, bending actuators are operated near the fundamental resonance frequency, due to the relatively low excitation voltage required to achieve large tip deflection [174], [181], [196], [197]. The first bending mode occurs at the fundamental resonance frequency, such that the peak stress is located near the clamped end of the bending actuator. Thus, a high potential for crack development is expected around 1 mm between the clamp and the protective coating on the piezoelectric layers.

Considering the inputs and outputs of the four-pole model for an unloaded bending actuator, the measurement chain for the accelerated lifetime tests is depicted in Figure 4.4.

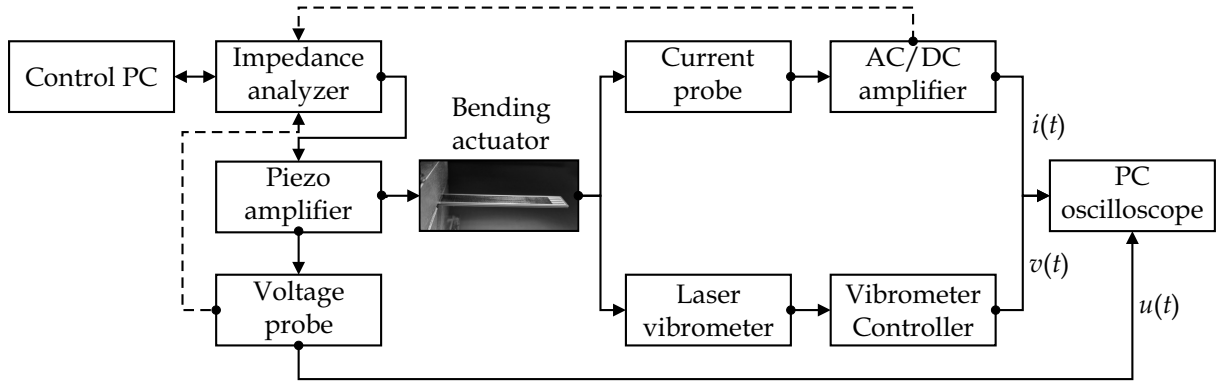


Figure 4.4: Measurement chain for the experiment with the bending actuator [175].

Through the coupling of a control PC and an impedance analyzer (HP 4192A LF), the amplitude and frequency of the input sinusoidal voltage can be specified. Since the bending actuator can theoretically be operated at higher voltages up to 230 V and the impedance analyzer has a maximum output voltage of approximately 2 V, a piezo amplifier was employed to amplify the voltage, as required by the bending actuator. Furthermore, due to the input voltage limitations of the impedance analyzer and the PC oscilloscope, a differential voltage probe (Testec TT- SI 9002) was employed to attenuate the amplified input voltage. During the experiments, the input voltage signal and the output signals, that is, the velocity at the tip of the bending actuator and the output current, were synchronously sampled every 10 s at a sampling rate of 30 kS/s and a measurement duration of 0.5 s. Specifically, the tip velocity was measured with a laser vibrometer (Polytec OFV 512) and corresponding controller (Polytec OFV 3000), while the current was acquired via a Tektronix A6312 current clamp and a corresponding AM 503B current probe amplifier. As the actuator's internal properties, such as stiffness and consequently the resonance frequency, can change over time due to degradation or loading conditions, a phase-locked loop (PLL) control was employed to track the frequency by monitoring the phase difference between the current and voltage signals, as described in [198]. In other words, before each experiment, the initial fundamental resonance frequency was determined with the impedance analyzer via a frequency sweep at a voltage of approximately 1 V. Furthermore, the frequency is adjusted during operation through the bidirectional communication between the impedance analyzer and the control PC, such that the phase difference be-

tween the current and voltage signal is constant and remain at the frequency close to the peak of the electrical admittance's magnitude [175].

4.2 Data understanding and preparation

Three set of preliminary experiments are considered to investigate the effects of varying operating conditions. The first set of preliminary experiments investigate the influence of different operating voltages on the characteristics of the bending actuators. The second set of experiments investigate the influence of different operating voltages on the condition monitoring signals. The last set of preliminary experiments investigate the influence of time-varying operating voltage in addition to a viscous fluid. With the findings from the preliminary experiments, several run-to-failure experiments were subsequently conducted under time-varying voltage amplitudes and a viscous fluid. In most cases, although the bending actuator is still functional, the EoL or failure of the bending actuator is defined here as the occurrence of a crack in at least a piezoelectric layer of the actuator, to avoid unreliable operation. During the experiments, the data acquisition was not automatically terminated in the event of a crack, such that the number of measurements does not necessarily indicate the life time of the bending actuators. An overview of the run-to-failure experiments investigated in the course of this work are listed in Tables 4.2 and 4.3. The experimental data PT01 through PT04 are collected under constant operating voltage conditions for the preliminary experiments. The experiments P01 through P08 consider time-varying voltage amplitudes sampled from a uniform distribution, while experiments P09 through P11 consider time-varying voltage amplitudes with two or three distinct voltage levels. Furthermore, experiment P10 und P11 also consider a viscous fluid.

Table 4.2: Overview of the experimental data from preliminary accelerated lifetime tests

Experiment	DC amplitude / V	AC amplitude / V	Effective length / mm	No. of measurements
PT01	20	20	38	268
PT02	20	20	39	4007
PT03	30	30	39	108
PT04	40	40	39	783

Table 4.3: Overview of the experimental data from accelerated lifetime tests

Experiment	DC amplitude / V	AC amplitude / V	Effective length / mm	No. of measurements
P01	30	[15, 25]	38	1809
P02	30	[15, 30]	38	2445
P03	30	[15, 30]	38	7573
P04	30	[15, 25]	38	641
P05	30	[15, 25]	38	526
P06	30	[15, 30]	38	1798
P07	30	[15, 30]	38	172
P08	30	[15, 30]	38	227
P09	30	{18, 20, 30}	38	285
P10 (with viscous fluid)	35	{20, 30}	38.5	503
P11 (with viscous fluid)	35	{20, 30}	38.5	406

4.2.1 Influence of the input voltage on the characteristics of the bending actuator

Bending actuators can be actuated with an higher driving levels to achieve higher deflections [199]. Thus, the influence of the actuating voltage variation on the characteristics of a bending actuator is considered here. Specifically, in one experiment, the AC voltage is varied, while no direct current (DC) bias is applied during the frequency sweep around the fundamental resonance peak, namely between 180 Hz and 270 Hz. The resulting magnitude and phase of the electrical admittance for three different arbitrary AC voltages, namely, at approximately 0.7 V, 4.4 V and 8.4 V, are depicted in the top and bottom sub-figures of Figure 4.5(a), respectively. The top sub-figure of Figure 4.5(a) shows that the maximum value of the electrical admittance's magnitude decreases with increasing AC voltage. The decrease in the maximum value of the electrical admittance's magnitude is also observed by Shaoping et al. [200], Priya et al. [201] and Lee et al. [202], but not by Uchino et al. [203], where the maximum value of the electrical admittance's magnitude contrarily increases. As also shown in the bottom sub-figure of Figure 4.5(a), the minimum value of the electrical admittance's phase increases. As can be deduced

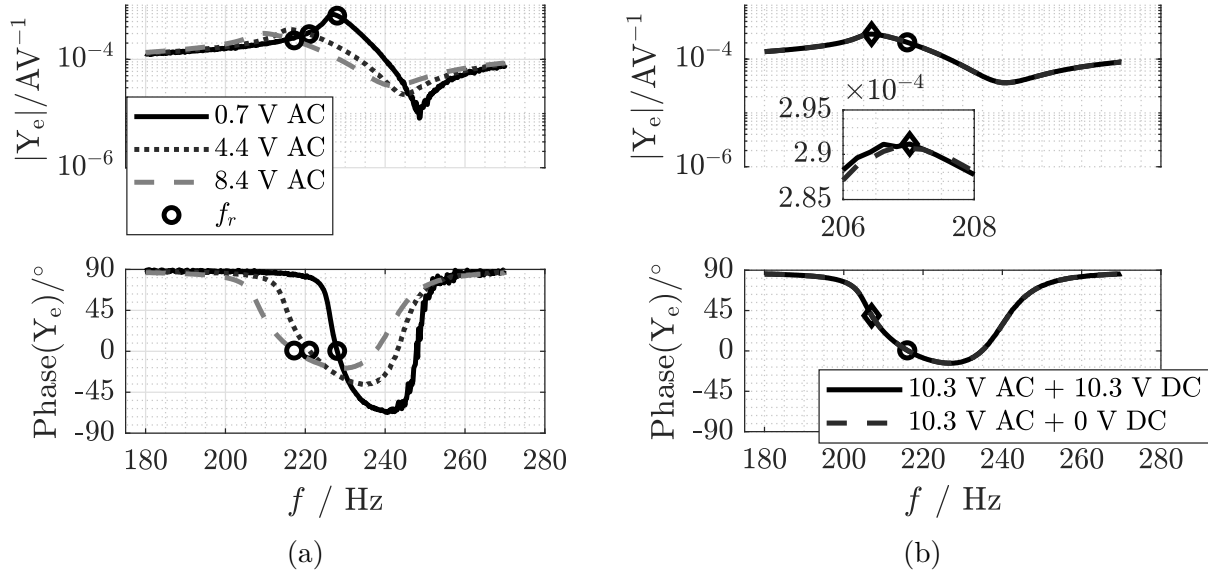


Figure 4.5: Amplitude (top) and phase (bottom) of the electrical admittance, under consideration of (a) different arbitrary AC voltages and without DC bias, and (b) constant AC voltage, with and without DC bias

from Figure 4.5(a), the resonance frequency also decreases with increasing AC voltage, as indicated by the circle marker. Although the decrease in resonance frequency could also result from internal and external characteristics of the bending actuator [174], [188], [204], the decrease in resonance frequency at higher AC voltage and the two preceding effects are supported by findings reported in the literature [200], [201], [202], [203], [205], [206] and are attributable to the softening non-linearity and internal non-linear loss of the bending actuator.

In the last preliminary experiment, the AC voltage is kept constant at approximately 10 V, while the magnitude and phase of the electrical admittance is measured at a DC bias of approximately 0 V and 10 V. Figure 4.5(b) shows that the DC bias variation has a negligible effect on the magnitude and phase of the electrical admittance even at a relatively high AC voltage, for the considered bending actuators. A close examination of both sub-figures of Figure 4.5(b) indicates that the admittance curve is further flattened and widened. Furthermore, the resonance frequency at a phase of 0° does not coincide with the peak of the electrical admittance's magnitude, which is achieved at a phase of approximately 42° . In summary, based on the considered voltage levels, the variation of the DC bias has a negligible influence on the characteristics of the investigated bending actuator, as opposed to the variation of the AC voltage that has a significant influence on the characteristics of the investigated bending actuator.

4.2.2 Influence of the input voltage on the condition monitoring signals

To investigate the influence of the input voltage amplitude on the measured output current i and velocity v at the tip of the bending actuator, two sets of run-to-failure experiments are considered. Specifically, the experiments are run till the current amplitude signifi-

cantly drops without a concurrent significant change in the voltage amplitude.

In the first set of experiment, the DC bias and the AC voltage amplitude are both kept constant at approximately 20 V during the entire experiment, while the current and velocity at the tip are measured over time for two bending actuators PT01 and PT02. The measured input voltage, the output current and velocity signals are exemplarily depicted in Figure 4.6. In the experiments, the frequency was adjusted, such that the respective phase difference between the current and voltage signal is constant and remain at the frequency close to the peak of the electrical admittance's magnitude. Furthermore, because the signals from both experiments are not synchronized, there is a phase shift in the sinusoidal waveform of the figures. Nonetheless, as observed in Figure 4.6(a), the signal mean is approximately 20 V, while the signal oscillates approximately between 0 V and 40 V, although the peak-to-peak amplitude of experiment PT02 is slightly above experiment PT01. This slight increase is also reflected in the measured output current and velocity signals as can be deduced from Figures 4.6(b) and 4.6(c), respectively. Thus, representing a direct relationship between the voltage input signal and the output signals. In both experiments, the maximum output current is approximately 5 mA, while the maximum output velocity is approximately 1 m s^{-1} .

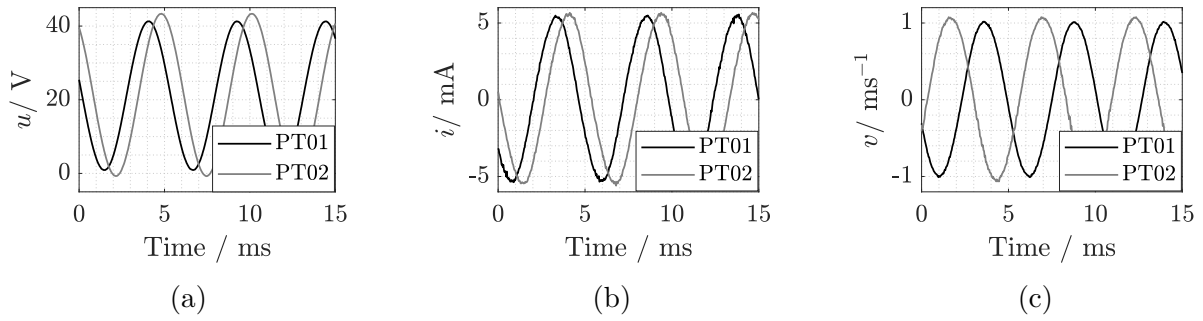


Figure 4.6: Comparison of the measured raw signals for two different bending actuators PT01 and PT02 exemplarily at about 3 min into the respective experiment: (a) input voltage signal, (b) output current signal and (c) output velocity signal

For further time-domain analysis, the RMS of the measured input voltage $\hat{u}$, consisting of the AC and DC signal components, the resonance frequency f_r , the RMS of the output current $\hat{i}$ and the RMS of the output velocity $\hat{v}$ are depicted in Figure 4.7 for both experiments PT01 and PT02 over their entire experiment duration. The top sub-figure of Figure 4.7(a) shows that the voltage RMS is almost constant at approximately 26.5 V, the resonance frequency is constant at about 189 Hz almost over the entire experiment duration, but decreases significantly with about 4 Hz shortly before the end of experiment PT01. Since there is no external influence during the experiment, it can safely be assumed that an internal change, such as a crack must have occurred in the bending actuator of experiment PT01. As presented in the bottom sub-figure of Figure 4.7(a), the velocity RMS is almost constant over the entire experiment duration, but slightly drops almost at the end of experiment PT01. As also presented in the bottom sub-figure of Figure 4.7(a), the current RMS slightly decreases over the entire experiment duration, and significantly drops at almost the end of experiment PT01, coinciding with the decrease in resonance frequency. As can be deduced from the top sub-figure of Figure 4.7(b), in experiment PT02, the voltage RMS is almost 1 V less than the voltage RMS of experiment PT01,

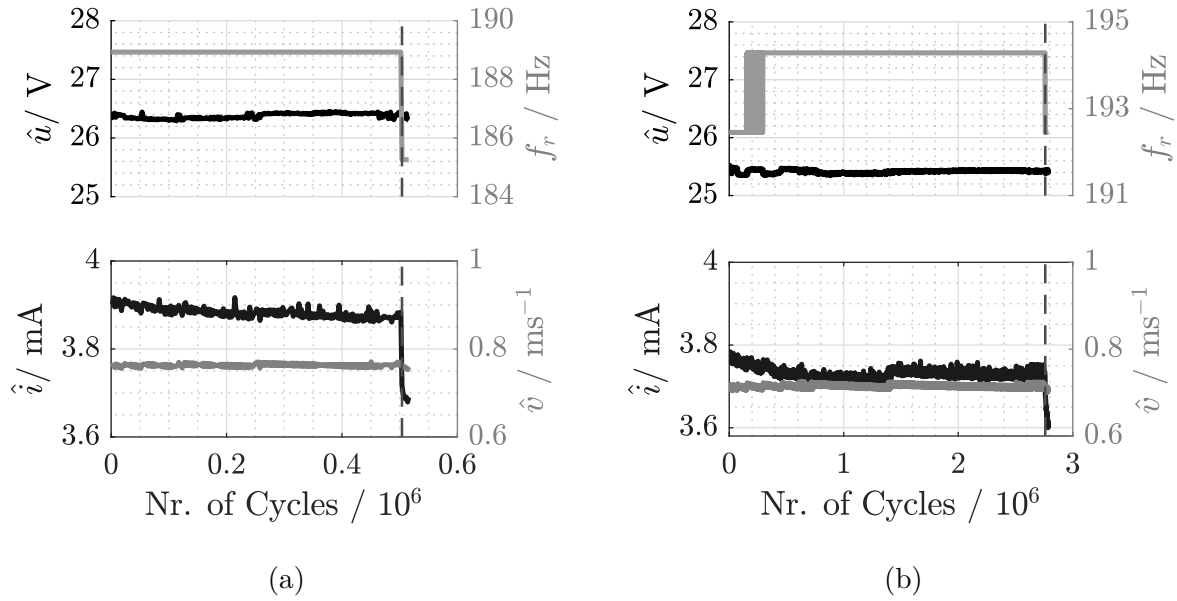


Figure 4.7: Comparison of the resonance frequency and RMS of the measured input and output signals for two bending actuators at a constant DC bias and AC voltage of approximately 20 V, respectively: (a) PT01 and (b) PT02

but the resonance frequency is almost 5 Hz more than experiment PT01, resulting from slight difference of the free length or internal characteristics of the bending actuators [207]. The fluctuation in the resonance frequency till about 300 000 cycles is attributed to the frequency control system. The bottom sub-figure of Figure 4.7(b) shows that the output velocity RMS is almost constant over the entire experiment duration, while the current RMS slightly declines over the entire experiment duration and significantly drops near the end of experiment PT02, similar to experiment PT01. However, although the output current RMS of both experiments PT01 and PT02 only slightly differ by about 0.1 mA, their lifetimes differ by a factor of approximately six. In summary, even at approximately the same input DC bias and AC voltage amplitude, the resonance frequencies and the lifetimes of two bending actuators can vary significantly. However, the decreasing trend of the current RMS tends to be consistent with a probable degradation of the bending actuators.

In the second set of experiment, the effect of an increased DC bias and AC voltage amplitude is investigated. Specifically, in experiment PT03, the DC bias is kept constant at 30 V and the AC voltage is also kept constant at 30 V during the entire experiment. In experiment PT04, the DC bias and AC voltage are kept constant at 40 V each, over the entire experiment duration. The top sub-figure of Figure 4.8(a) shows that the voltage RMS is almost constant at about 37 V over the entire measurement duration for experiment PT03. The resonance frequency is constant at 187 Hz almost over the entire experiment duration, but drops significantly to about 183 Hz at 200 000 cycles, as a result of a possible internal anomaly of the bending actuator. As shown in the bottom sub-figure of Figure 4.8(a), the output velocity RMS fluctuates around a constant value of approximately 0.9 ms^{-1} over the entire experiment duration, while the current RMS slightly declines from approximately 5.4 mA to 5.2 mA in about 200 000 cycles and thereafter significantly drops to approximately 5 mA coinciding with the decrease in frequency and consistent

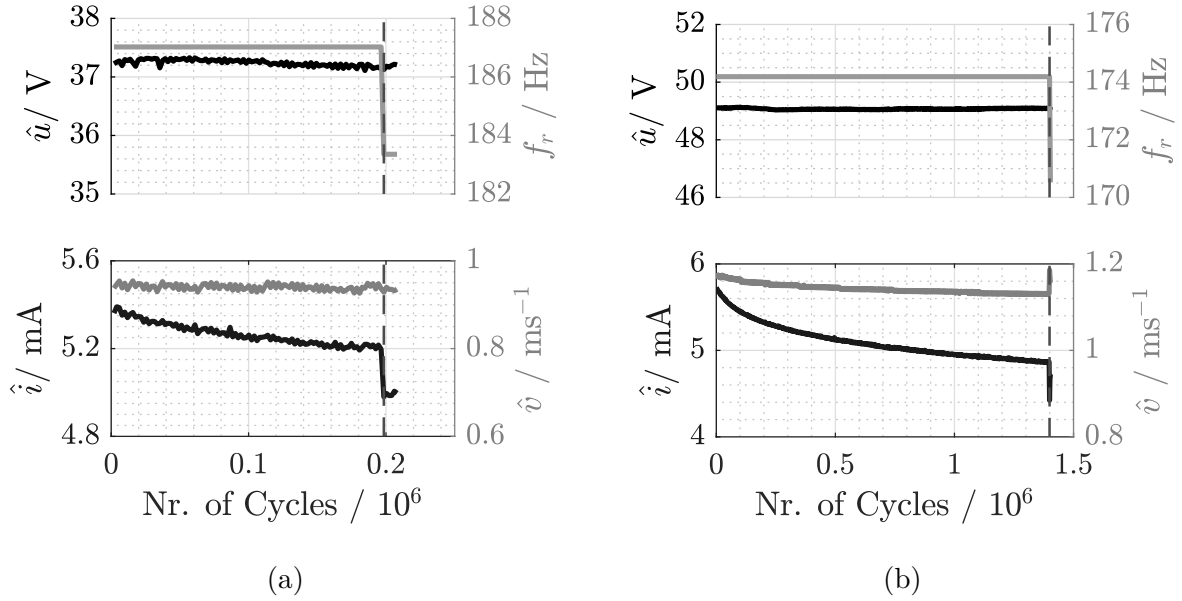


Figure 4.8: Comparison of the measured input and output signals for two different bending actuators at different constant DC bias and AC voltage: (a) 30 V DC bias + 30 V AC (PT03) and (b) 40 V DC bias + 40 V AC (PT04)

with the previous set of experiments PT01 and PT02. In experiment PT04, the voltage RMS is almost 10 V more than in experiment PT03, while the resonance frequency is about 13 Hz less than in experiment PT03, as can be deduced from the top sub-figure of Figure 4.8(b). Thus, it can be inferred that an increased voltage results in a decreased resonance frequency, which is coherent with the previous investigation in section 4.2.1. The bottom sub-figure of Figure 4.8(b) shows that the velocity RMS slightly decreases with about 0.05 ms^{-1} and increases to almost its initial value at about 1.4 million cycles. Contrarily, in experiment PT01, the velocity RMS simultaneously drops with the decrease in frequency. Such contradiction is accountable to the fact that the bending actuator is not completely failed, but anomalous. The bottom sub-figure of Figure of 4.8(b) also shows that the current RMS is approximately 0.4 mA more than in experiment PT03 resulting from the increased voltage RMS. Furthermore, the decline in the current RMS over the entire experiment duration is more pronounced than the decline in experiment PT03. Similar to the previous experiments, the current RMS significantly drops at about 1.4 million cycles simultaneously with the decrease in frequency.

The resulting images from post-experimental microscopic crack analysis of experiments PT01, PT03 and PT04 are exemplarily presented in Figure 4.9, where the arrows indicate the crack location in the top view, in the region shown in Figure 4.3. Although not visible from the images, all cracks occurred in the piezoelectric layer through the entire width in close proximity to the clamped region. The crack width is wider in experiment PT04 than in experiment PT01 and PT03, due to the relatively higher voltage RMS. However, the difference in the crack width between experiments PT01 and PT03 is not very pronounced, although the voltage RMS differs by almost 10 V.

From the experiments, it can be summarized that an increased voltage results in a decreased resonance frequency, an increased velocity and increased current. However, in a run-to-failure experiment, an increase in voltage of 10 V does not necessarily lead to a

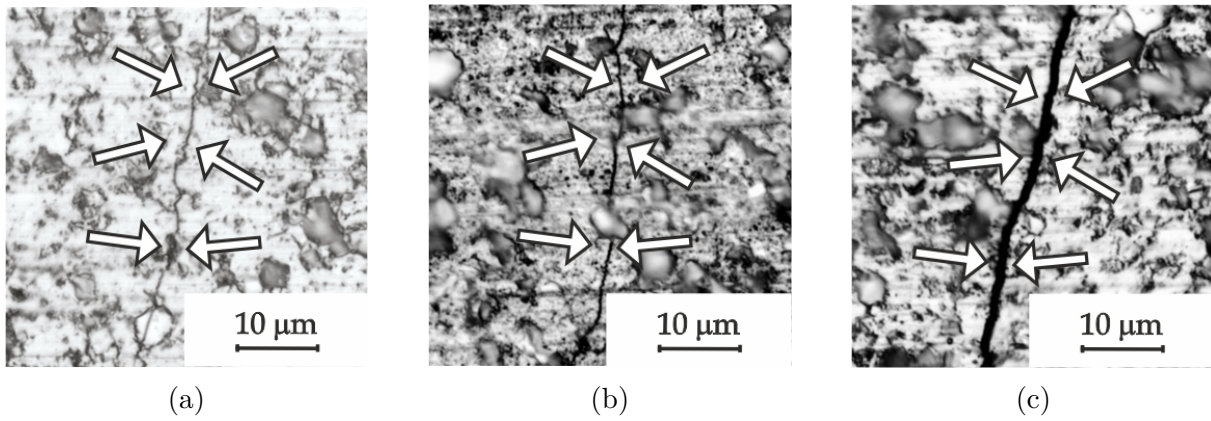


Figure 4.9: Comparison of the crack widths resulting from the experiments with three bending actuators subjected to different constant DC bias and AC voltage: (a) 20 V DC bias + 20 V AC (PT01), (b) 30 V DC bias + 30 V AC (PT03) and (c) 40 V DC bias + 40 V AC (PT04)

crack occurring earlier in the piezoelectric layer of the bending actuator, but the crack width becomes widened, due to the larger deflection resulting from the increased velocity. Even at similar operating voltage conditions, the variation in lifetimes of the bending actuators can be due to manufacturing variability, which might occur during producing or bonding of the individual materials in the different layers [208]. The velocity signal is not further considered, due to the inconsistency in relation to a possible anomaly in the bending actuator.

4.2.3 Influence of time-varying voltage amplitude and a viscous fluid [175]

A run-to-failure experiment P06 with time-varying voltage amplitude without a viscous fluid and an experiment P10 with a viscous fluid is exemplarily considered to further investigate the influence of time-varying operating conditions on the monitored condition monitoring signals. In experiment P06, the DC bias is constant at 30 V, while the AC voltage is randomly sampled from a uniform distribution within the interval [15, 25] V, utilizing the setup for the unloaded bending actuator, as previously described. In some studies [209], [210], [211], [212], it has been shown that damping increases and the resonance frequency of a bending actuator decreases when the actuator is immersed within viscous fluid. Thus, in experiment P10, the bending actuator is immersed in a viscous fluid. During the experiment, the bending actuator is unloaded and actuated with a constant DC bias of 25 V and a two-level AC voltage of approximately 25 V and 35 V at the first resonance frequency to accelerate degradation. To investigate the influence of the increased damping on the fault detection capability, the actuator is immersed in a glass beaker with a mixture of several Reely silicone oils with a resulting viscosity between 4 Pa s and 6 Pa s at room temperature [175], after the occurrence of a crack and during operation. The actuator is partially submerged in three different levels to investigate the influence of different damping levels. In the first level, the resulting depth of actuator within the viscous fluid is approximately 11 mm, as shown in Figure 4.10(a) and labeled as position A in Figure 4.10(b). The actuator is then further immersed in the viscous liquid in two steps, as labeled in positions B and C in Figure 4.10(b). The maximum

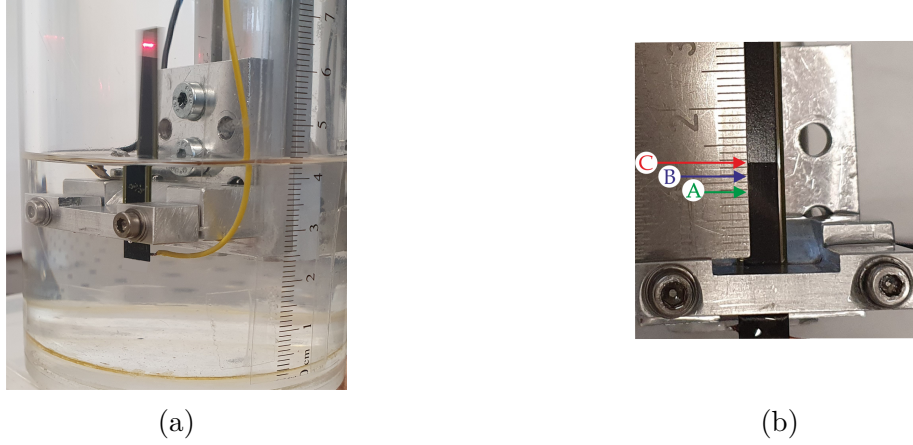


Figure 4.10: Setup for experiment P10 showing the (a) bending actuator immersed in silicone fluid and (b) different positions, when immersed within silicone fluid [175].

depth of actuator within the viscous fluid is approximately 15 mm [175].

The time-domain analysis results of experiments P06 and P10 are depicted in Figures 4.11(a) and 4.11(b), respectively. As shown in the top sub-figure of Figure 4.11(a), generally, an increase in the input voltage amplitude consequently results in a decreased resonance frequency. However, at about 650 000 cycles, even with no significant increase in the input voltage RMS, the resonance frequency significantly decreases, possibly due to a reduced stiffness. Concurrently, the current RMS also slightly decreases at about 650 000 cycles, as shown in the bottom sub-figure of Figure 4.11(a). Contrarily, the velocity RMS does not show any significant change at about 650 000 cycles.

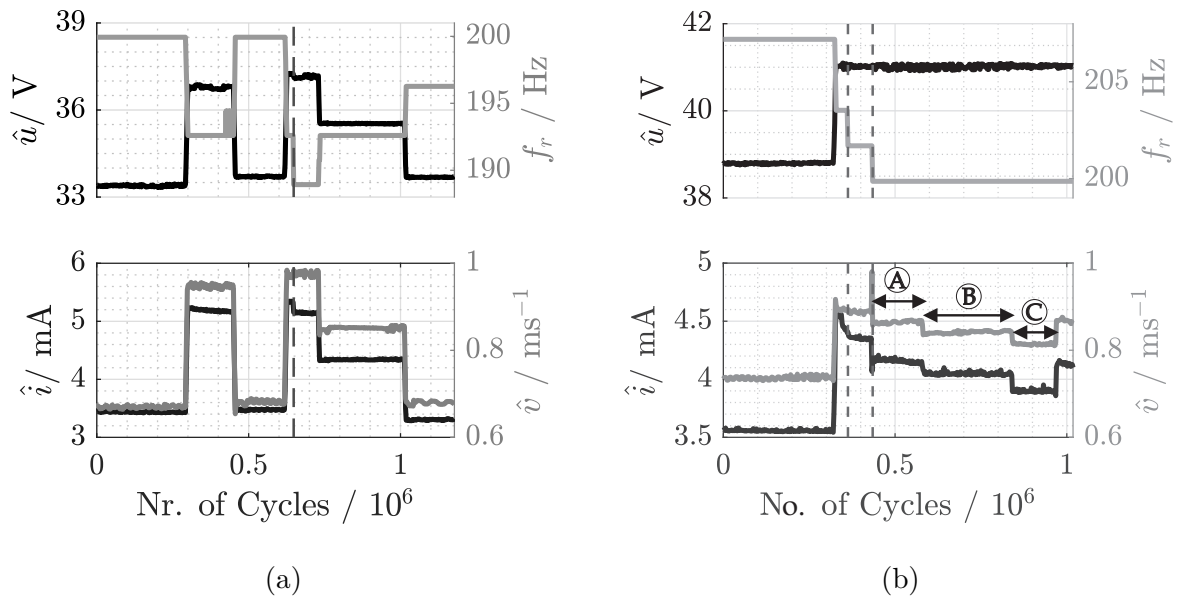


Figure 4.11: Resulting analysis results of experiments considering (a) time-varying voltage amplitude in experiment P06 and (b) additionally considering a viscous fluid in experiment P10

However, an increased voltage amplitude generally leads to an increased current and velocity amplitude, while a decreased voltage amplitude also results in a decreased current and velocity amplitude, as presented in the bottom sub-figure of Figure 4.11(a). In experiment P10, an increased voltage amplitude generally leads to a decreased resonance frequency, an increased current amplitude and an increased velocity amplitude, as shown in Figure 4.11(b). However, as presented in the top sub-figure of Figure 4.11(b), at approximately 360 000 cycles and 440 000 cycles, the frequency significantly decreases, although there is no concurrent significant increase in the voltage amplitude. The decrease in frequency at about 360 000 cycles is probably attributable to a change in the internal characteristics of the bending actuator, such as a decrease in stiffness, while the decrease in frequency at about 440 000 cycles coincides with immersing the actuator in the viscous liquid, which is known to lead to increased damping and consequently to decreased frequency. Furthermore, as illustrated in the bottom sub-figure of Figure 4.11(b), the current amplitude further decreases when the bending actuator is further immersed in the viscous fluid, as indicated by the positions B and C in Figure 4.10(b). The current amplitude further increases at about 970 000 cycles, because the actuator is lifted out of the viscous liquid.

In summary, a sudden decrease in the current amplitude and a simultaneous decrease in the resonance frequency suggest a change in the internal or external characteristics of the bending actuator.

4.3 Fault detection approaches

In the following, different approaches are investigated to detect a fault in the bending actuator, resulting from the mode of operation and loading situation. Precisely, the fault is in form of a hairline crack in at least a piezoelectric ceramic layer near the clamped region. Regarding the conclusions drawn from the preliminary experiments, the current signals are further considered, but not the velocity signals.

4.3.1 Approach A: Signal analysis techniques [175]

The time-frequency domain analysis method toward fault detection is investigated here. To this end, the spectrograms of the current signals of experiments P06 and P10 are exemplarily depicted in Figures 4.12(a) and 4.12(b), respectively. As described in the previous section, the actuating frequency is not constant, but varies according to the load. Thus, in both figures, the frequencies on the y-axis are normalized to the actuating frequency, that is, the first resonance frequency, to reveal the harmonics. As shown in both figures, the first harmonic, which corresponds to the first resonance frequency, is dominant over the entire measurement duration. Furthermore, higher harmonics are also present, exemplarily up to the seventh order. Harmonics up to the seventh order are considered here, because the effects of harmonics higher than the seventh order were not significantly relevant for the considered application example and analysis techniques. As can be inferred from Figure 4.12(a), before approximately 650 000 cycles, the intensity of the third harmonic is more dominant than the other higher harmonics. However, after 650 000 cycles, even and other odd harmonics are also visible in the spectrogram. Although not easily discernible from Figure 4.12(b), the intensity of the odd harmonics, that is, the third, fifth and seventh harmonics are more dominant after approximately 360 000 cycles. According to Solodov et al. [213], the clapping of crack surfaces typically results in odd

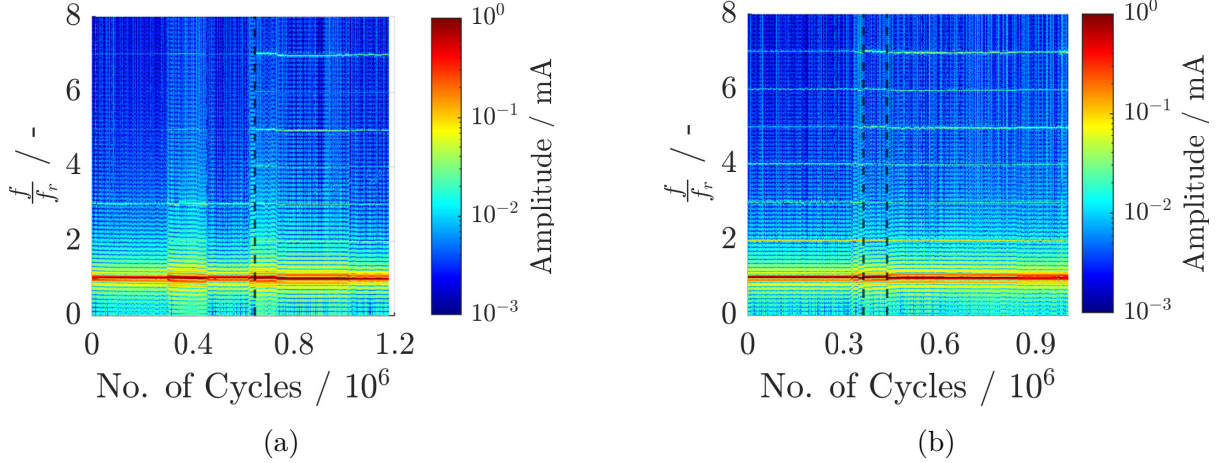


Figure 4.12: Time-frequency analysis of the current signals for experiments (a) P06 and (b) P10.

and even harmonics, while predominantly odd harmonics are attributed to friction non-linearity of the crack surfaces [175]. Thus, the higher intensity of the odd harmonics after 650 000 cycles and 360 000 cycles for experiments P06 and P10, respectively, can be attributed to non-linearity of possible crack surfaces. In general, higher harmonics result from signal distortion, which can be quantified individually or globally, according to the DIN EN 50160 standard [214]. Individually, harmonics can be quantified via the normalized or relative amplitudes of the harmonics, that is, the ratio of a higher harmonic amplitude to the fundamental amplitude [214], as expressed in equation 4.2.

$$\tilde{I}_n = \frac{\hat{I}_n}{\hat{I}_1}, \quad (4.2)$$

where $\tilde{I}_n$ is the normalized amplitude value of the n th harmonic current signal, $\hat{I}_n$ is the amplitude value of the n th harmonic content and $\hat{I}_1$ is the amplitude value of the fundamental component [175]. Globally, harmonics can be quantified via the total harmonic distortion (THD), as expressed in equation 4.3 [214], [215].

$$\text{THD} = \frac{\sqrt{\sum_{n=2}^N (\hat{I}_n)^2}}{\hat{I}_1} = \sqrt{\sum_{n=2}^N \left(\frac{\hat{I}_n}{\hat{I}_1} \right)^2} = \sqrt{\sum_{n=2}^N (\tilde{I}_n)^2}. \quad (4.3)$$

The resulting normalized amplitudes up to the seventh harmonic for experiments P06 and P10 are depicted in Figures 4.13(a) and 4.13(b), respectively. Specifically, the normalized amplitude values of the even harmonics, that is, harmonics of order 2, 4, and 6 are shown in the top sub-figures, while the normalized amplitude values of the odd harmonics of order 3, 5, and 7 are shown in the bottom sub-figures. As can be observed from the top sub-figure of Figure 4.13(a), although the normalized amplitude value of the second harmonic $\tilde{I}_2$ does not show any noticeable trend before 650 000 cycles, it increases significantly at about 650 000 cycles and remains almost twice the initial value till the experiment was terminated. On the other hand, the normalized amplitude value of the fourth harmonic $\tilde{I}_4$ only slightly increases at about 650 000 cycles, while the normalized amplitude value of

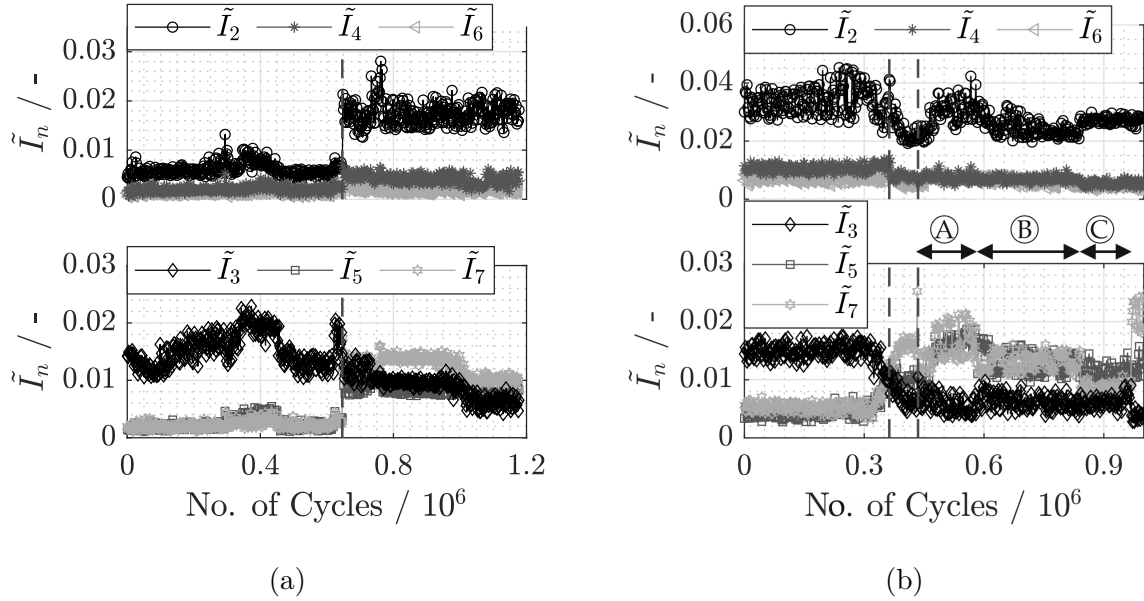


Figure 4.13: Normalized amplitudes at even and odd harmonics up to the seventh harmonic for experiments (a) P06 and (b) P10

the sixth harmonic $\tilde{I}_6$ does not show any significant trend over the entire measurement duration. Similarly for experiment P10, $\tilde{I}_6$ does not also show any significant trend over the entire measurement duration, as can be observed from the top sub-figure of Figure 4.13(b). Contrarily to experiment P06, $\tilde{I}_2$ and $\tilde{I}_4$ slightly decreases at about 360 000 cycles and do not show any significant trend over the entire measurement duration for experiment P10. As evident from the bottom sub-figure of Figure 4.13(a), at about 650 000 cycles, the normalized amplitude value of the third harmonic $\tilde{I}_3$ slightly decreases, while the normalized amplitude value of the fifth and seventh harmonics, $\tilde{I}_5$ and $\tilde{I}_7$ simultaneously significantly increases. As also observed from the bottom sub-figure of Figure 4.13(b), $\tilde{I}_3$ also slightly decreases, while $\tilde{I}_5$ and $\tilde{I}_7$ simultaneously significantly increases at about 360 000 cycles. As the bending actuator is being immersed in the viscous fluid labeled at position A at about 430 000 cycles, $\tilde{I}_5$ and $\tilde{I}_7$ abruptly decreases due to damping, while $\tilde{I}_3$ remains almost constant. As the actuator is further immersed in positions B and C at about 580 000 cycles and 840 000 cycles, respectively, $\tilde{I}_5$ and $\tilde{I}_7$ slightly decreases, while $\tilde{I}_3$ contrarily slightly increases. Finally, at about 970 000 cycles, as the actuator is lifted out of the viscous liquid, $\tilde{I}_7$ significantly increases, $\tilde{I}_5$ remains almost constant, while $\tilde{I}_3$ concurrently decreases.

In summary, the evaluation of the normalized amplitude values of the odd harmonics $\tilde{I}_3$, $\tilde{I}_5$ and $\tilde{I}_7$ presents a possibility to detect an anomaly and isolate a possible crack in the bending actuator.

For the global quantification of the harmonics, the resulting THD values for both experiments P06 and P10 are depicted in the top sub-figures of Figures 4.14(a) and 4.14(b), respectively. As shown in the top sub-figure of Figures 4.14(a), the THD value increases to almost twice its initial value at about 650 000 cycles and slightly decreases towards the end of the measurement, corresponding to the decreased voltage amplitude. Contrarily, the THD value does not show any significant trend for experiment P10, as can be observed from the top sub-figure of Figures 4.14(b). Thus, because the THD values are not

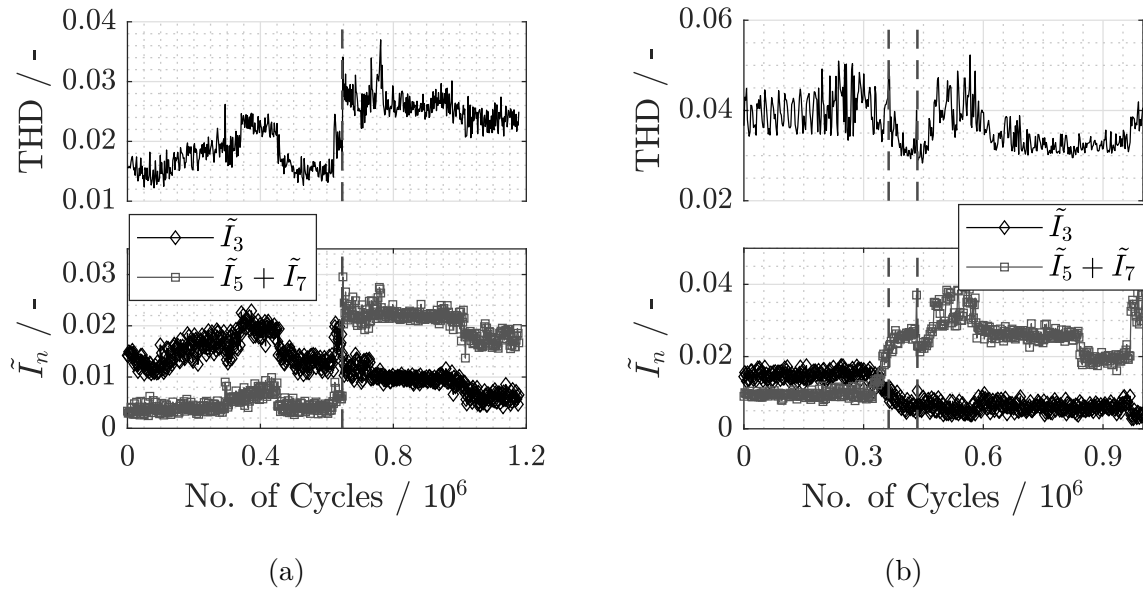


Figure 4.14: Comparison of the THD and normalized amplitudes at odd harmonics up to the seventh harmonic for experiments (a) P06 and (b) P10

consistent with the degradation of the bending actuators, a crack in the bending actuator is isolated by considering the interaction between the normalized amplitudes of the odd harmonics. Specifically, a crack in at least a piezoelectric layer is isolated by evaluating $\tilde{I}_3$ and the sum of $\tilde{I}_5$ and $\tilde{I}_7$. As shown in the bottom sub-figures of Figure 4.14, the value of $\tilde{I}_3$ is greater than the sum of $\tilde{I}_5$ and $\tilde{I}_7$, before a crack, while the value of $\tilde{I}_3$ is less than the sum of $\tilde{I}_5$ and $\tilde{I}_7$ after a crack.

The procedure for detecting an anomalous behavior and isolating a crack in a bending actuator is depicted in Figure 4.15. In summary, an anomaly is detected in the bending actuator, when the current amplitude $\hat{i}$ abruptly drops and the resonance frequency f_r simultaneously decreases. However, since such an anomaly could result from external damping or from the effect of crack, the amplitudes of the odd harmonics are further evaluated to isolate a crack.

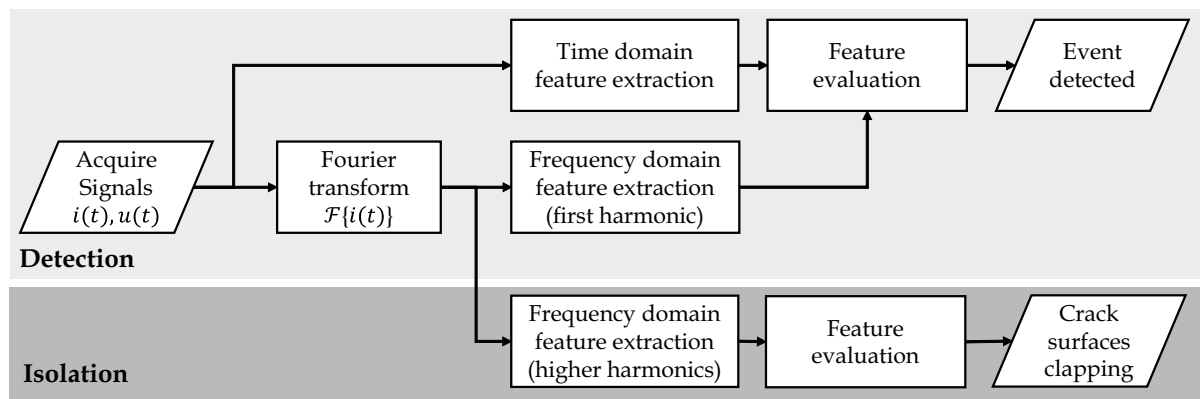


Figure 4.15: Fault detection and isolation of a piezoelectric bending actuator [175].

4.3.2 Approach B: Condition-invariant feature engineering

Similar to the procedure described in [137], classical time-domain features, features from the HCTSA package and features from the EMG toolbox are extracted from the acquired current signals per experiment. Subsequently, the feature selection method, monotonicity is employed to identify condition-invariant features. Although unique top-ranked feature that are invariant to the time-varying operating voltage could not be identified per experiment with monotonicity, a eight-element subset of the ranked features could be identified via explorative analysis. The eight features are from the HCTSA package and allow a distinct clustering of each experiment into two groups of healthy and faulty instances. Exemplarily, the `DN_OutlierInclude_abs_001.stdrfexpr2` feature is derived from distributional outliers statistics, while the `PP_Compare_resample_1_2.kscn_peaksepy` is derived from time-series pre-processing properties [216]. Figure 4.16(a) exemplarily shows the `DN_OutlierInclude_abs_001.stdrfexpr2` for the two experiments P06 and P10.

As observed in the figure, the feature value is noisy over the entire measurement duration and decreases at about 360 000 cycles for experiment P10, but significantly increases for experiment P06 at about 650 000 cycles, both corresponding to the onset of an anomaly. However, in combination with the other feature subset, it allows for a clustering into groups 0 and 1, namely healthy and unhealthy instances, as exemplarily shown for experiment P10 in Figure 4.16(b), where `DN_OutlierInclude_abs_001.stdrfexpr2` is plotted against `PP_Compare_resample_1_2.kscn_peaksepy`.

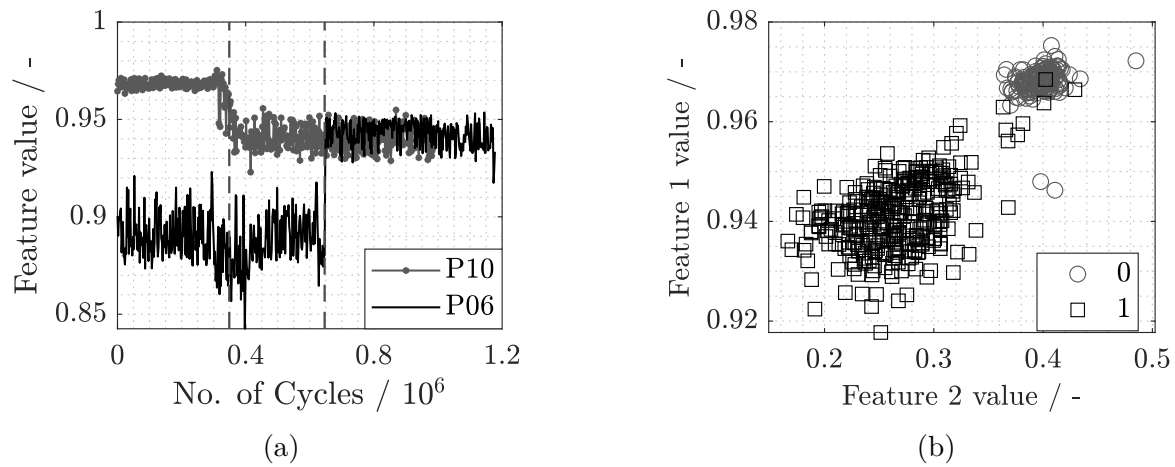


Figure 4.16: Comparison of a subset of the selected features for experiments P06 and P10: (a) `DN_OutlierInclude_abs_001.stdrfexpr2` and (b) `DN_OutlierInclude_abs_001.stdrfexpr2` (Feature 1) and `PP_Compare_resample_1_2.kscn_peaksepy` (Feature 2)

4.3.3 Approach C: Condition-centered data or feature scaling

The current signal has been previously identified to reflect the degradation of the bending actuator. Thus, in the following, features derived from the current signal are scaled by features from the time-varying voltage signal over the entire duration of the experiment. Specifically, since the output current signal is directly proportional to the voltage signal, the ratio of the RMS current signal, $\hat{i}$ to the RMS voltage signal, $\hat{u}$ is calculated. The resulting electrical admittance $\tilde{Y}_e$ for experiment P06 is shown in Figure 4.17(a). As seen

in the figure, the influence of the time-varying voltage is very dominant, even after scaling the RMS current signal, $\hat{i}$ by the RMS voltage signal, $\hat{u}$. Alternatively, the fundamental signal components are utilized to compute the electrical admittance $\tilde{Y}_e$, that is, the ratio of $\hat{I}_1$ to $\hat{V}_1$. As can be inferred from Figure 4.17(b), the influence of the time-varying voltage is almost eliminated for both experiments. However, to detect a fault in the experiments, an expert-based threshold or further processing is required, because the range of values for healthy and faulty instances are different in both experiments.

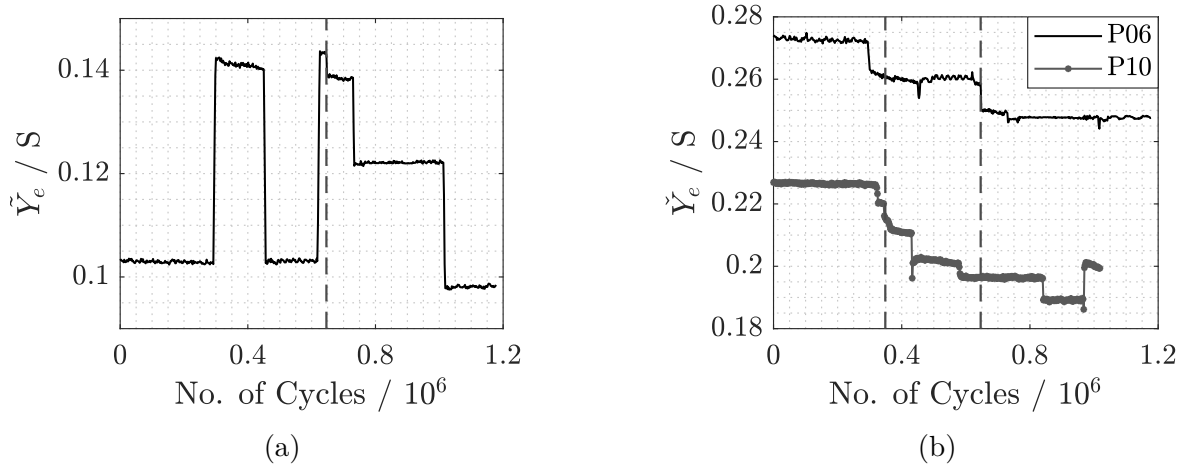


Figure 4.17: Feature scaling based on: (a) RMS values for experiment P06 and (b) fundamental signal components for experiments P06 and P10

4.3.4 Approach D: Condition-specific feature standardization and feature segmentation

Before feature standardization and segmentation, features derived from the input voltage signal, as the sole operating condition variable, are considered as inputs to the k-means clustering algorithm to identify distinct operating regimes. Specifically, the RMS, variance and standard deviation of the input voltage signal are employed as inputs to the k-means algorithm. Furthermore, the Davies-Bouldin index is employed to derive the optimum number of clusters. The Davies-Bouldin index evaluates cluster performance via intra-cluster and inter-cluster distances and results in metric values in the range $[0, \infty]$, where a value close to zero indicates well-separated clusters [217], [218]. The resulting cluster evaluation with the Davies-Bouldin index for twenty clusters is depicted in the top sub-figure of Figure 4.18(a), exemplarily for experiment P06. As shown in the figure, the minimum value of the Davies-Bouldin index is ten, which indicates the optimal number of clusters. However, the knee point is located at five clusters, as also shown in the top sub-figure of Figure 4.18(a), which indicates a trade-off between optimality and complexity. The resulting clusters from the k-means algorithms with five initial clusters and the input voltage features are exemplarily depicted in the bottom sub-figure of Figure 4.18(a). As seen in the figure, there is not an abrupt transition between two voltage levels, but intermediate voltage values occur, which can be considered as outliers. Thus, the five initial clusters are justified over ten initial clusters, because the outliers do not form individual clusters. The standardization of the current feature based on the mean and standard deviation per voltage cluster indices according to Equation 3.1 in Section 3.2.4

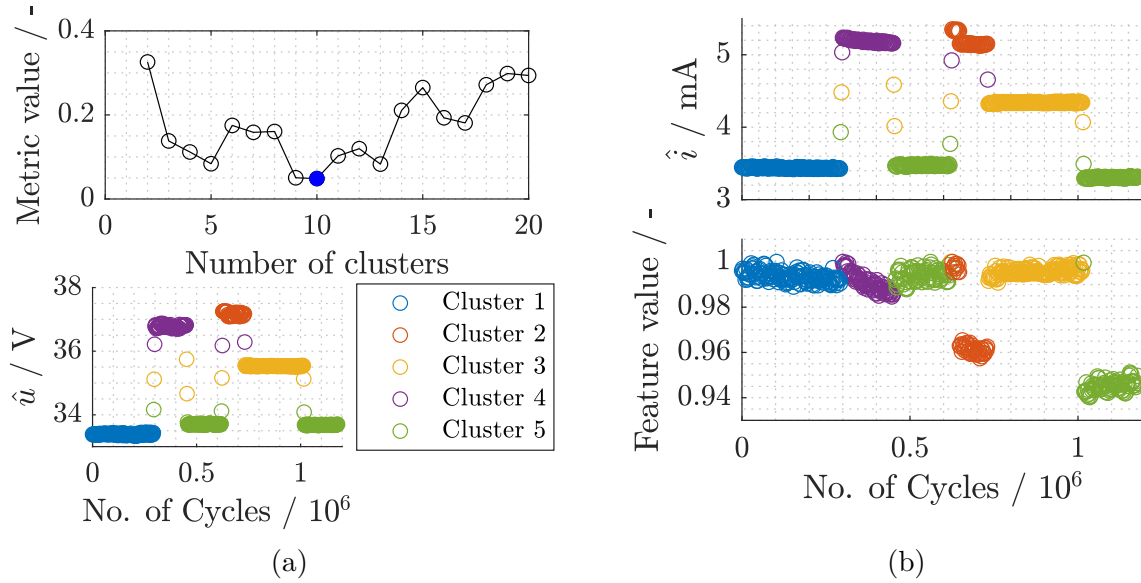


Figure 4.18: Cluster analysis of the input voltage and output current for experiment P06: (a) cluster evaluation of the input voltage and (b) corresponding clustering of the output current without (top) and with feature scaling per voltage cluster (bottom).

did not yield any significant result. Alternatively, the current RMS is segmented per voltage cluster indices, as illustrated in the top sub-figure of Figure 4.18(b). Furthermore, each current feature segment is scaled by its maximum value, after outlier removal via the generalized extreme studentized deviate test [219], [220]. The resulting scaled current RMS per voltage cluster is shown in the bottom sub-figure of Figure 4.18(b). As can be inferred from the figure, an anomaly can be detected for the experiment P06, when the derived feature value crosses a value of 0.97. However, the derived feature value does not reflect an anomaly in the third cluster segment, even after an anomaly had occurred in the second cluster segment, because there is no reference value for the third cluster segment in the healthy state. Hence, this approach is not further considered here.

4.3.5 Approach E: Embedding in the modeling phase

Fault detection in the piezoelectric bending actuator can be formulated as a binary clustering task involving healthy instances before the occurrence of a crack and faulty instances with the occurrence of a crack in at least a piezoelectric layer of the bending actuator. To this end, features derived from the current signal are considered as inputs to the k-means clustering algorithm. The resulting clusters based on the RMS, variance and standard deviation of the current signal are shown in the top sub-figure of Figure 4.19(a), exemplarily for experiment P06. As shown in the figure, the clusters do not allow the distinction between healthy and faulty instances. Even the consideration of the voltage features as further input to the k-means algorithm does not improve on the clustering results, as shown in bottom sub-figure of Figure 4.19(a). The Davies-Bouldin index values for 20 clusters is depicted in the top sub-figure of Figure 4.19(b) to determine the optimum number of clusters, while the resulting clusters based on the optimum number of three

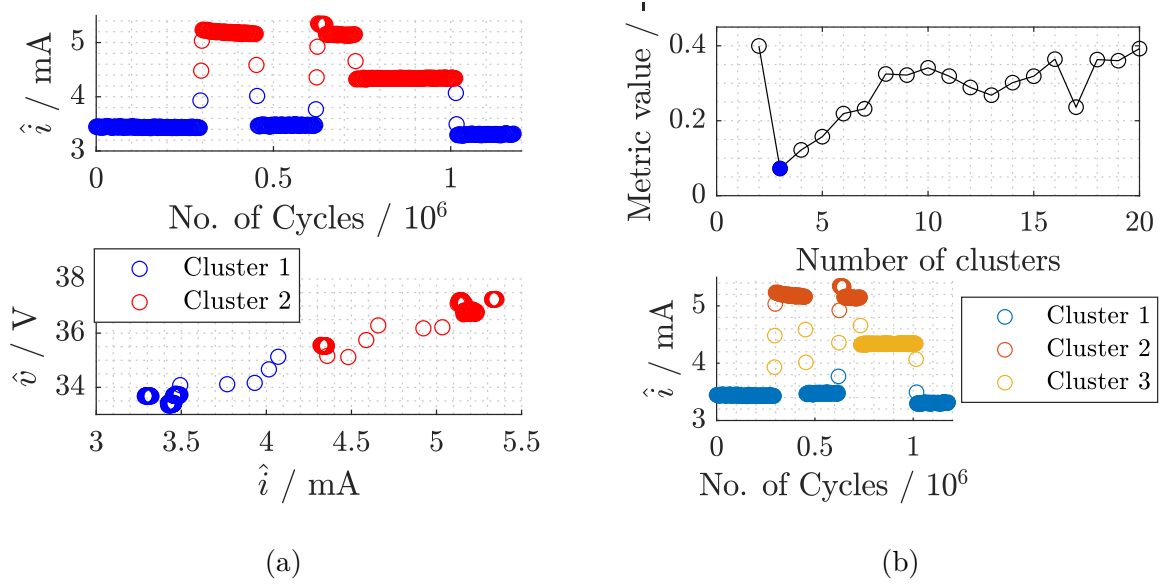


Figure 4.19: Cluster analysis of the current signal for experiment P06: (a) with features from the current signal (top) and in addition to the voltage signal (bottom). (b) Cluster metric evaluation of the current signal and resulting clusters based on three clusters

clusters is depicted in the bottom sub-figure of Figure 4.19(b).

The bottom sub-figure of Figure 4.19(b) shows that the resulting clusters only allow a distinction between the current levels of low, medium and high current and not a distinction between healthy and faulty instances. Alternatively to the binary clustering task, a regression task can be formulated that involves predicting the fault-free current signal. To this end, features from the input voltage signal, such as the RMS, variance and standard deviation can be employed as inputs to an arbitrary regression model, such as a regression tree to predict the RMS of the fault-free current signal. As shown in the top sub-figure of Figure 4.20(a) for experiment P06, the training input data, namely the voltage RMS is exemplarily depicted alongside the fault-free current RMS as the training target. The model prediction result and the true RMS of the current signal is shown in the bottom sub-figure of Figure 4.20(a) over the entire measurement duration. As can be inferred from the figure, the model precisely estimates the true RMS of the fault-free current signal. However, at about 650 000 cycles onward, the model estimations significantly deviate from the true current RMS, which indicates a possible faulty state in the bending actuator. Under consideration of the pre-trained model of experiment P06, the model estimation results for experiment P10 are represented in Figure 4.20(b). As shown in the top sub-figure of Figure 4.20(b), the model estimations significantly deviates from the true current RMS over the entire measurement duration, making it difficult to discern if the deviation is fault-related or not. However, even in the consideration of a physics-based model, as investigated by Bender [181], such deviation between the model and the measured quantity also occurred in the healthy state. Therefore a metric, such as an error metric or ratio between the true and estimated values must be considered to determine when the bending actuator becomes faulty. Furthermore, when considering an error metric between the estimated current RMS value $\hat{p}\hat{i}$ and the true current RMS value

$\hat{a}\hat{i}$, such as the absolute error $|\hat{p}\hat{i} - \hat{a}\hat{i}|$, as exemplarily shown in the bottom sub-figure of Figure 4.20(b), a threshold is necessary for fault detection.

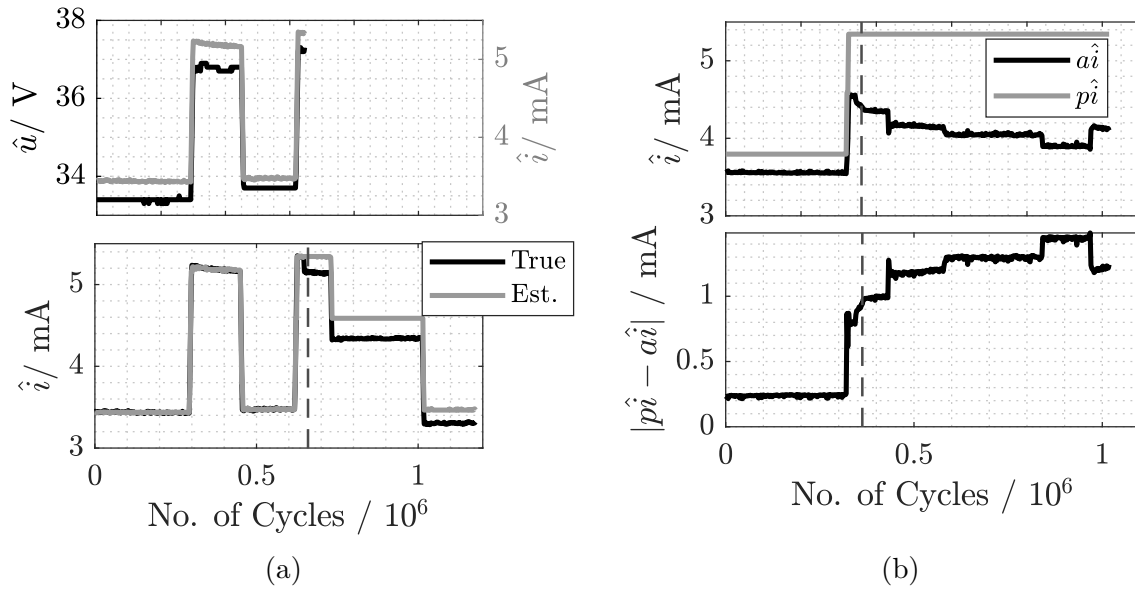


Figure 4.20: Prediction of the current RMS based on features from the input voltage signal: (a) voltage RMS and the current RMS before a crack, as part of the training data and target data, respectively (top) and prediction of the current RMS for the entire measurement duration (bottom) for experiment P06. (b) Prediction of the current RMS for the entire measurement duration for experiment P10 and respective absolute error $|\hat{p}\hat{i} - \hat{a}\hat{i}|$, where $\hat{p}\hat{i}$ and $\hat{a}\hat{i}$ denote the estimated and the true current RMS values, respectively

4.4 Performance evaluation of the fault detection approaches

To prevent interrupting the experiments during operation, microscopic crack analysis of the bending actuators was conducted after the experiments, which revealed an hairline crack in at least a piezoelectric layer. Furthermore, since it is not known in advance precisely when and where the hairline crack in the piezoelectric layers will occur, it was infeasible to automatically obtain a very high-resolution image shortly before the crack. Therefore, the ground truth is unavailable. Consequently, the evaluation of the presented fault detection methods by the confusion matrix or metrics derived thereof are not feasible. However, for the purpose of comparison, the transition points, that is, the measurement indices at crack and corresponding number of cycles (Cyc.) in million cycles are presented in Table 4.4 for experiments P01 through P11. The resulting transition points determined with approach A, and specifically referring to the procedure illustrated in Figure 4.15, are considered as the reference for comparison. In the following, a binary k-means algorithm is employed to determine the initial partition point between healthy and faulty instances. In approach B, the input data to the k-means algorithm are eight condition-invariant features, as described in Section 4.3.2, while the input data in approach C is the scaled

fundamental current signal $\hat{I}_1$ with respect to the fundamental voltage signal $\hat{V}_1$. In approach E, the RMS, variance and standard deviation of the voltage signal are considered in addition to similar features of the current signal as input data to the k-means algorithm.

Generally, as observed from Table 4.4, the life times of the bending actuators subjected to similar operating conditions are different, due to material and manufacturing variability. For example, although experiments P04 and P05 were subjected to similar operating conditions, a crack occurred in experiment P05 almost 5.6×10^5 cycles later than in experiment P04. A crack also occurred later in experiment P06 than in experiment P08, although also subjected to similar operating conditions. As can be inferred from Table 4.4, approach E yielded too early transition points and underperformed, as compared to the other approaches. The results discussed in section 4.3.5 share light on the poor performance of approach E, where the resulting clusters only allow the distinguishability between the current signal levels and not a distinction between healthy and faulty instances. Approach C leads to better transition points, as compared to approach E in most of the experiments. However, it underperforms approach B for experiments P01, P02, P05, P06, P07, and P08, where approach B provides similar transition points to the reference approach A. Furthermore, approach B only minimally deviates from approach A, for the other experiments.

Thus, condition-invariant features derived from the current signals can be considered as an alternative to approach A in distinguishing between a healthy and unhealthy actuator, even when operating under time-varying conditions.

Table 4.4: Comparison of the transition times for experiments P01 through P11 with the results obtained via approach A as the reference approach

Experiment	Approach A		Approach B		Approach C		Approach E	
	Transition point		Transition point		Transition point		Transition point	
	Index / -	Cyc. / 10^6	Index / -	Cyc. / 10^6	Index / -	Cyc. / 10^6	Index / -	Cyc. / 10^6
P01	325	0.69	325	0.69	1795	3.73	201	0.43
P02	2416	5.16	2416	5.16	2444	5.22	141	0.30
P03	679	1.37	687	1.39	679	1.37	153	0.31
P04	78	0.16	80	0.17	78	0.16	57	0.12
P05	360	0.72	360	0.72	93	0.19	91	0.19
P06	312	0.65	312	0.65	218	0.45	141	0.29
P07	115	0.24	116	0.24	114	0.24	88	0.18
P08	81	0.16	80	0.16	94	0.19	76	0.15
P09	276	0.55	170	0.35	170	0.35	169	0.35
P10	168	0.36	159	0.34	200	0.43	150	0.32
P11	200	0.43	143	0.31	195	0.42	143	0.31

4.5 Prognostics approaches

Several approaches are also considered here to estimate the RUL of the bending actuators subjected to time-varying voltage amplitudes. As previously stated, the EoL is defined here as the occurrence of a crack in at least a piezoelectric layer of the actuator to avoid unreliable operation. Thus, the RUL prediction task here involves estimating the cycle remaining before the occurrence of a crack in at least a piezoelectric layer of the actuator. The resulting features derived with approach A, that is signal analysis techniques are not considered here. Furthermore, condition-specific feature standardization and feature segmentation from approach D, both do not directly represent the monotonic degradation of the examined bending actuators and hence are not feasible here. Thus, three approaches of considering the operating condition information are investigated, while employing several machine learning methods to map selected features to the RUL. analogous to the supporting paper [169]. Concretely, approach B solely utilizes features derived from the current signal as an input to the machine learning methods. Approach C considers the current signal features scaled by the voltage signal features, while approach E considers voltage signal features, in addition to the current signal features. The general procedure in estimating the RUL is illustrated in Figure 4.21.

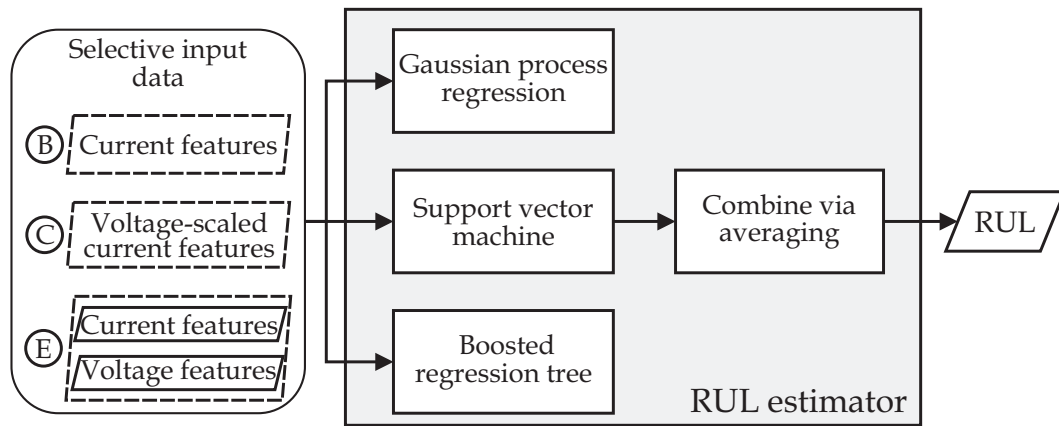


Figure 4.21: General procedure towards estimating the RUL of bending actuators subjected to time-varying voltage amplitudes

Depending on the selected approach, the input features are subsequently preprocessed and passed on to the respective machine learning method, namely an exponential Gaussian process regression (GPR), a Gaussian SVM and a boosted regression tree. During training, the target is the RUL in million cycles. Furthermore, the leave-one-out strategy is employed, where all but one experiment is employed for training and the remaining is employed as test data. For example, the RUL estimation results for experiment P06 is obtained by excluding the experiment P06 during training. The predicted RUL from the individual algorithms are subsequently combined via averaging to produce an ensemble RUL. In the following, the three approaches are elaborated upon, while exemplarily considering experiment P06.

4.5.1 Approach B: Current features as sole inputs to machine learning regression models

In section 4.3, features extracted from the current signal were found to facilitate the detection of a crack in the bending actuators, even under time-varying operating voltage conditions. Although the extracted features are applicable here, the selected features are not directly applicable, because features for a binary decision, if the bending actuator has a crack or not, might not be sufficient for a continuous decision of the healthiness of the bending actuator. Thus, a set of ten features are reselected with monotonicity for the RUL estimation task. The selected features include the band power delta in the frequency range [200, 300] Hz computed utilizing the EEG feature extraction toolbox, features from symbolic transformations and discrete wavelet transform coefficients of the Daubechies wavelet calculated with the HCTSA toolbox. Due to the time-varying voltage amplitude conditions and the corresponding abrupt jumps in the feature values, a linear model is fit to each feature segment, which is demarcated by the change points in the voltage signal, as depicted by the smoothed solid line in Figure 4.22(a). Under the assumption that the bending actuator is initially healthy, the feature values are also subsequently scaled by the initial feature value, as depicted in the right y-axis of Figure 4.22(a). The smoothed and scaled features are separately employed to train an exponential GPR model, a Gaussian SVM and a boosted regression tree, each.

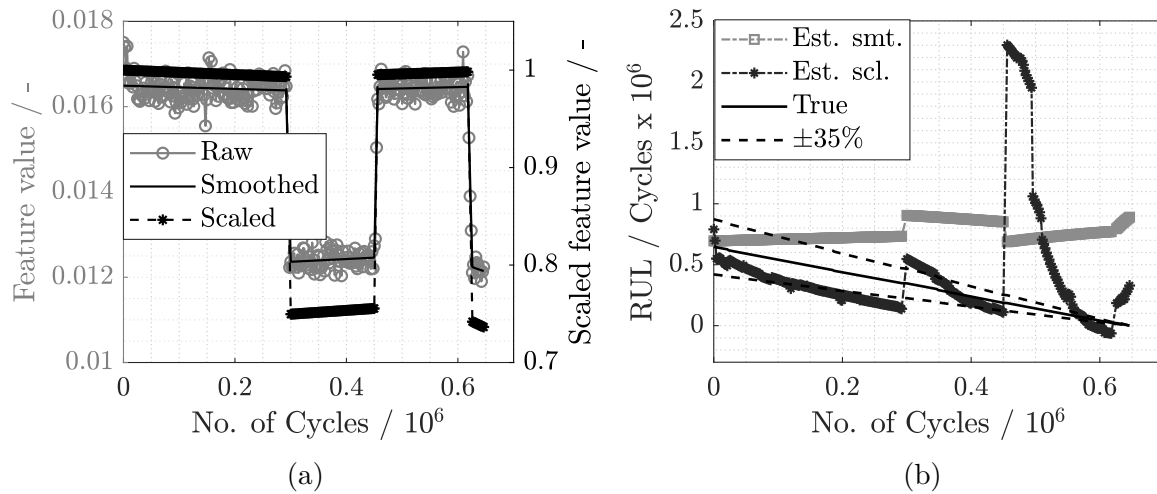


Figure 4.22: Current features as input data and corresponding estimated RUL for experiment P06. (a) Comparison of two feature post-processing steps to the unprocessed current feature. (b) Comparison of the RUL resulting from the smoothed and scaled current features

The final RUL obtained by averaging the individual predictions of the three models is depicted in Figure 4.22(b), exemplarily for experiment P06. Figure 4.22(b) shows that the mean RUL obtained with the smoothed and scaled feature values both reflect the effects of the operating condition, such that the abrupt change in the feature values, as can be observed from Figure 4.22(a) also result in abrupt change in the RUL estimations. Figure 4.22(b) also shows that the mean RUL obtained with the smoothed feature values is mostly overestimated. Furthermore, although the estimated RUL remains within the error-margin till about 113 000 cycles, it is almost constant. Contrarily, the mean

RUL obtained with the scaled feature values linearly decreases, although mostly underestimated. Thus, since overestimation can lead to adverse or extreme consequences than underestimation, only the mean RUL obtained with the scaled feature values are further considered.

4.5.2 Approach C: Voltage-scaled current features as inputs to machine learning regression models

In this approach, the current features from section 4.5.1 are scaled by the standard deviation of the voltage signal per experiment, to compensate for the effects of the time-varying voltage amplitudes. In a further post-processing step, the resulting feature values are then scaled by the initial feature values per experiment. The feature values before and after the post-processing step are depicted in Figure 4.23(a), exemplarily for experiment P06. Figure 4.23(a) shows that the feature trend is intensified via the post-processing step. Thus, similar to the preceding approach, the post-processed features are then employed to train an exponential GPR model, a Gaussian SVM and a boosted regression tree. The final RUL obtained by averaging the individual predictions of the three algorithms is depicted in Figure 4.23(b) exemplarily for experiment P06. As depicted in Figure 4.23(b), the estimated RUL generally outperforms the RUL estimated by the foregoing two approaches, especially towards the EoL. However, at certain time periods, such as from the BoL till approximately 166 000 cycles, the estimated RUL outperforms approach E, but underperforms the approach A, described in sections 4.5.1 and 4.5.3, respectively.

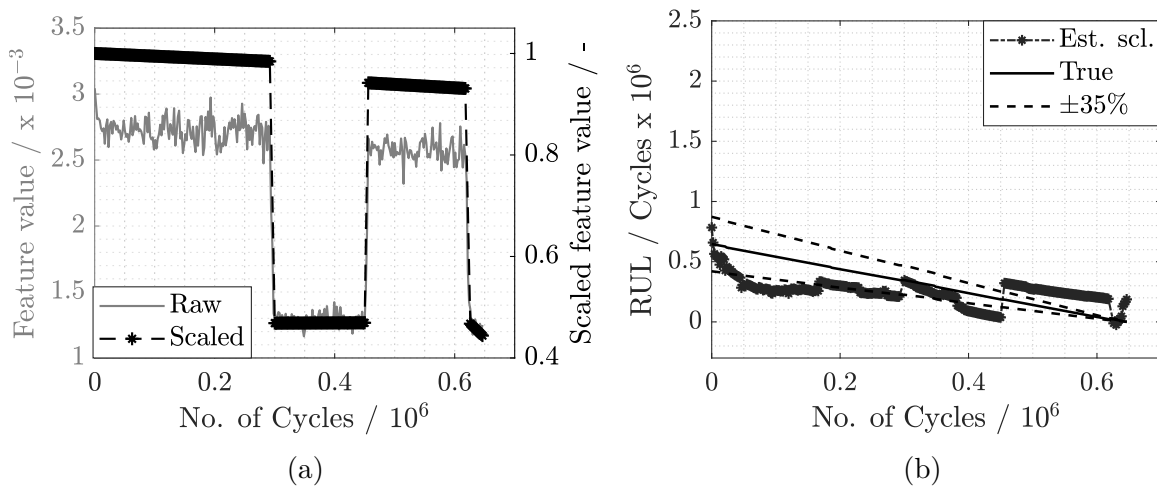


Figure 4.23: Voltage-scaled current features as input to estimate the RUL for experiment P06. (a) Comparison of the unprocessed and the scaled feature value and (b) corresponding RUL based on the scaled feature value

4.5.3 Approach E: Voltage features in addition to current features as inputs to machine learning regression models

Since the current signal is directly influenced by the voltage signal, the RMS of the voltage signal is considered in addition to the current features from section 4.5.1 as a further input during the training of the machine learning models, in this approach. The current features are scaled by the initial feature values, as described in the previous section, while the voltage RMS is min-max normalized to the $[0, 1]$ interval. The resulting feature values after the application of the scaling factor is shown alongside the voltage RMS in Figure 4.24(a), exemplarily for experiment P06. As shown in the right y-axis of Figure 4.24(a), the feature values of the test data lie within the range of the training data. During the training of the individual models, there was no significant effect of the min-max normalized voltage signal on the implementation of the SVM and GPR models, because the model predictors are thereafter standardized in both models. However, a positive effect of the min-max normalization could be observed during the training of the boosted regression tree. The final RUL obtained by averaging the individual predictions of the three algorithms is depicted in Figure 4.24(b), exemplarily for experiment P06. As compared to the mean estimated RUL of the first approach, the prediction accuracy slightly deteriorates, such that the mean estimated RUL for this approach is further underestimated till approximately 300 000 cycles, as shown in Figure 4.24(b). Furthermore, the mean estimated RUL is overestimated approximately between 300 000 cycles and 450 000 cycles. Thereafter, although the mean estimated RUL is further overestimated, it outperforms the model estimation of the approach A till approximately 620 000 cycles.

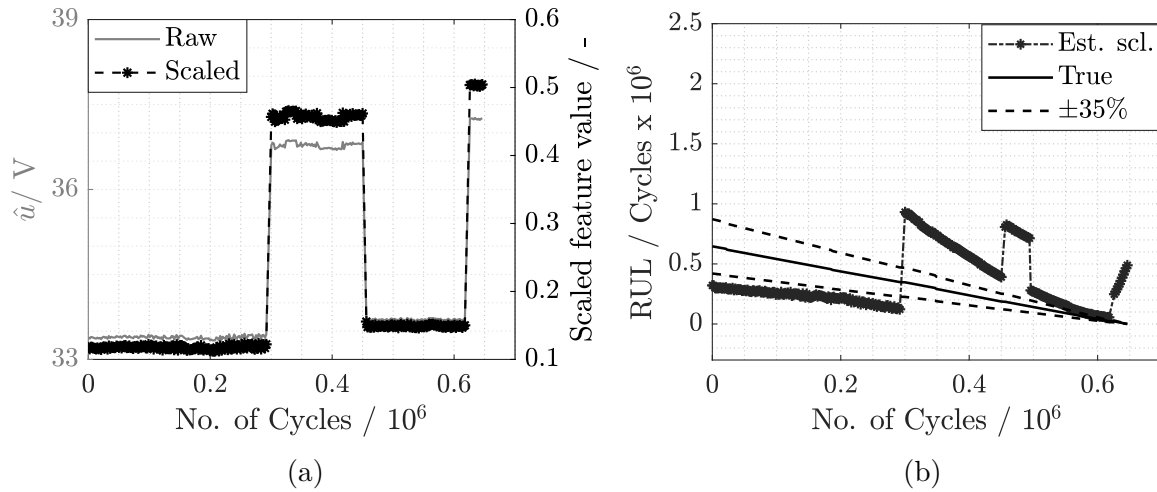


Figure 4.24: Voltage feature as an additional input to estimate the RUL for experiment P06. (a) Comparison of raw and scaled voltage RMS and (b) corresponding RUL based on the scaled current and voltage features

4.6 Performance evaluation of the prognostics approaches

The wMAPE and the nRMSE metrics are employed to evaluate the performance of the proposed prognostic approaches, because the metrics allow for a quantitative comparison over several datasets. The resulting wMAPE values for the experiments P01 through P11 are presented in Table 4.5 for the three approaches, while the resulting nRMSE values are presented in Table 2, in the Appendix, for conciseness. However, the evaluation results based on the nRMSE support the evaluation results of the wMAPE. In Table 4.5, approach B represents the ensemble RUL model utilizing solely the features derived from the current signal. Approach C considers voltage-scaled current features as input data to estimate the RUL, while approach E incorporates voltage features in addition to the current features. Approach B is considered as the baseline in the following, because the same features are utilized in approaches C and E.

As shown in Table 4.5, with the exception of experiments P01, P02, P06, P09, P10, approach C consistently underperforms approach B. However, it overperforms the other approaches for experiments P02 and P06. Contrarily, approach E demonstrates superior performance over approach C in more than 63% of the considered experiments. Furthermore, approach E demonstrates superior performance over or equivalent performance to approach B in more than 80% of the considered experiments.

From the evaluation results it can be summarized that the incorporation of the voltage features in addition to condition-invariant features derived from the current signals yield to improved performance in estimating the RUL of the actuators, even under time-varying operating conditions.

Table 4.5: Performance evaluation with the wMAPE metric for the three described approaches

Experiment	wMAPE		
	Approach B	Approach C	Approach E
P01	5.81	5.57	4.17
P02	0.92	0.91	0.92
P03	0.63	0.94	1.05
P04	1.88	3.32	3.45
P05	3.82	4.88	1.33
P06	0.87	0.47	0.87
P07	19.29	20.22	11.57
P08	9.29	10.60	8.59
P09	1.22	1.17	0.74
P10	2.22	1.14	1.06
P11	1.49	2.53	1.08

5 Fault detection and prognostics of ball bearings

This chapter presents experimental data of ball bearings subjected to time-varying loads and speeds as an application example to demonstrate the effectiveness of the proposed fault detection and prognostics methodology.

Bearings are components of rotating machinery that allow the relative positioning of other machine elements, such as the bearing housing and shaft, while providing support and rotational freedom. Furthermore, bearings help to minimize rotational friction and to transmit axial and radial load from the rotating component to the supporting structure [77], [221], [222]. Thus, the stable operation of rotating machinery depends on the quality and performance of bearings [222]. Bearings are broadly categorized into sliding and rolling element bearings, depending on the underlying friction between the contact components [222]. Depending on the load direction, rolling element bearings are further categorized into axial and thrust bearings. Additionally, rolling element bearings are generally classified into ball bearings, cylindrical roller bearings, needle roller bearings and tapered roller bearings, depending on the shape of rolling element [222].

Although this application example focuses on ball bearings, rolling element bearings generally composes of a cage, an inner ring, an outer ring, rolling elements, and optionally seals as shown in Figure 5.1(a). The relevant dimensions of the ball bearing considered in this application example is exemplarily depicted in Figure 5.1(b), where d denotes the inner diameter, D represents the outer diameter, B implies the width, D_T denotes the rolling circle diameter and D_W represents the rolling element diameter. Around the world, approximately 10 billion bearings are manufactured yearly. While 90% of the bearings remain intact during the lifetime of the machinery, about 9.5% of the bearings are replaced before their actual EoL, to prevent failure, and roughly 0.5% of the bearings, that is, approximately 50 million bearings are replaced due to failure [223]. Approximately 40% to 50% of all motor failures are attributed to such unplanned bearing failure [224], which can lead to human and economic losses and also to unplanned downtime of the entire machinery.

According to the ISO 281 standard [226], the life of a rolling element bearing is defined as the number of revolutions that a bearing can endure, before fatigue initiation in at least one of the bearing components, such as a spall in the inner raceway, becomes evident. Under consideration of high quality manufacturing, constant and similar operating conditions, the ISO 281 standard [226] provides a statistical-based approach to calculate the life of a rolling element bearing, namely the basic rating life. The basic rating life or L_{10} of a bearing is the reliability that 90% of a group of identical bearings subjected to similar operating conditions will reach or exceed the specified life. The influence of contamination, lubrication, wear, corrosion and extreme operating conditions are not directly covered by the basic rating life. However, several modifications have been made to consider such influences, such as the so-called modified rating life. The modified rating life allows for the specification of the reliability, for example L_{50} for 50% reliability instead

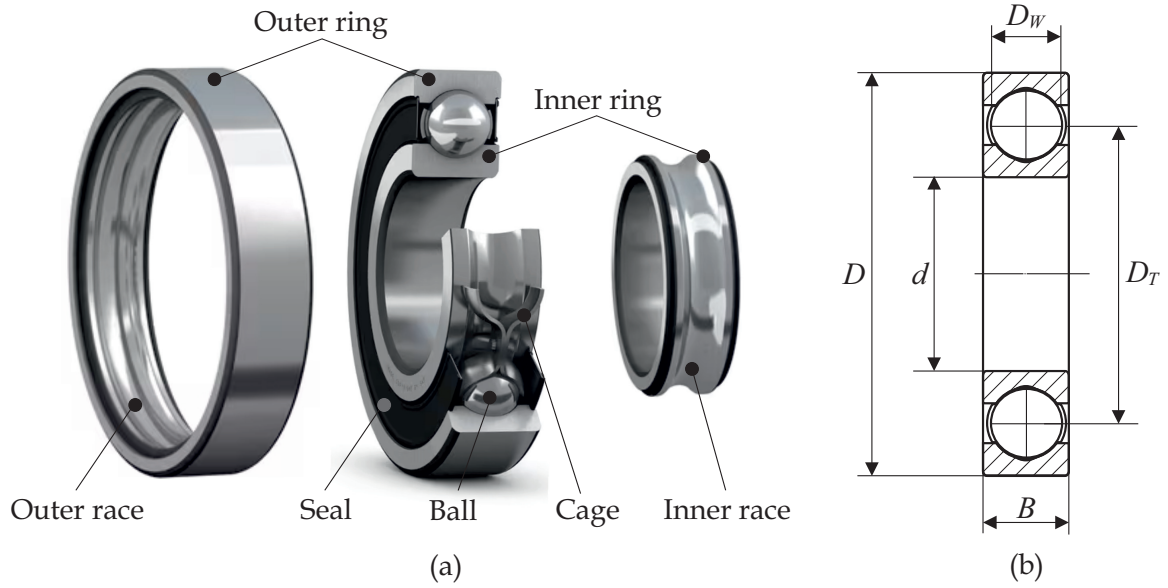


Figure 5.1: Rolling element bearing: (a) typical components of a ball bearing adapted from SKF [225] and (b) its dimensions

of 90%. Furthermore, operational factors, such as contamination levels and lubrication conditions, are also specifiable. The modified rating life L_{nm} , in millions of revolutions, is defined in equation 5.1.

$$L_{nm} = a_1 a_{ISO} L_{10} = a_1 a_{ISO} \left(\frac{C}{P} \right)^p, \quad (5.1)$$

where a_1 and a_{ISO} denote the modification factors for reliability and operational factors, such as the lubrication condition, respectively [226]. C represents the bearing-specific basic dynamic radial load rating, P represents the dynamic equivalent load and the bearing-dependent exponent p , $p = 3$ for ball bearings [226]. Theoretically a bearing can attain an infinite life, under favorable factors, such as cleanliness, operating conditions and lubrication. Because such favorable factors are not always attainable in real applications, where technical systems need to operate under harsh conditions, this work only covers run-to-failure experiments with a pronounced fault in at least one of the bearing components, such as on the rolling element, inner ring or outer ring.

Typical bearing failure modes, failure causes and monitoring techniques are presented in the subsequent section, with a special focus on rolling element bearings. Subsequently, the fault detection and prognostics methodology are evaluated through the application example.

5.1 Technical understanding

This section covers typical failure modes of ball bearings, techniques for bearing condition monitoring, test rig design and experimental setup.

5.1.1 Ball bearings failure modes

According to the ISO 15243 standard [227], the failure modes and failure mechanisms or causes of rolling element bearings are not easily distinguishable. Hence, based upon the features visible on the surfaces of bearing components, bearing failure is primarily classified in rolling contact fatigue, wear, corrosion, electrical erosion, plastic deformation, cracking, and fracture, under consideration of bearings with steel rolling elements and rings [227].

Rolling contact fatigue results from repeated stress changes in the contact zones of the rolling element and raceways and manifests as visible changes in the microstructure or macrostructure of rolling element and inner/outer raceways [227]. Microstructural changes are classified under subsurface initiated fatigue, which occurs around the maximum shear stress zone underneath the surface and typically results to microcracks initiated beneath the surface [223]. The depth of the microcracks is dependent on the operating temperature, applied load, the cleanliness and microstructure of the underlying material and the material itself [223], [227]. Macrostructural changes are classified under surface initiated fatigue resulting from plastic deformation of rolling contact surface asperities [223], [227]. Surface initiated fatigue is mostly attributed to inadequate or improper lubrication and can often result to surface microcracks and microspalls or pitting [223], [227].

Wear is the removal of material from the contact surfaces primarily as a result of hard particles or primarily as a result of insufficient lubrication, namely abrasive or adhesive wear, respectively [227]. Abrasive wear is typically attributed to low viscosity of the lubricant, small hard particles in the lubricant, and a combination of inadequate lubrication, extreme low speeds and high loads [223]. Although dull surfaces is peculiar to abrasive wear, shining surfaces can also occur resulting from polishing effect of tiny particles [223], [227]. Adhesive wear is mostly attributed to inadequate lubrication of the surfaces in contact and is characterized by smearing, which involves the material transfer from one surface to another [223], [227].

Corrosion occurs on bearing components as a consequence of chemical or electro-chemical reaction [227]. When the chemical reaction is activated by liquids, such as water or corrosive agents, moisture corrosion occurs on the components' surface, which further leads to corrosion pits and surface spalls [227]. However, when the chemical reaction is activated due to friction resulting from relative sliding of contact surfaces, then frictional corrosion occurs [227]. Frictional corrosion resulting from loose bearing fit is categorized under fretting corrosion, while frictional corrosion resulting from excessive vibration typically at standstill is categorized under false brinelling [223], [227].

Electrical erosion is attributed to electric current flow causing material removal or changes in the microstructure of the contact surfaces [227]. Electrical erosion occurs either as excessive current erosion, when the electric current passing through the contact surfaces is excessive, or as current leakage erosion, when the electric current intensity is extremely low. The material removal or craters in excessive current erosion is more pronounced and loosely positioned than in current leakage erosion. Furthermore, burns or discoloration are consequences of electrical erosion [223], [227].

Plastic deformation in form of overload deformation is a consequence of static overload, while particles-induced deformation occurs as a consequence of foreign particles between the surfaces in contact [227]. Overload deformation can also be caused by poor handling, improper mounting or shock loads, which results to true brinelling or flutes, that is, shallow indentations on bearing raceways, cages or rolling elements. Particles-induced

deformation occur as indentation on bearing components caused by the over-rolling of hard or soft particles between the raceways and rolling elements. Such indentation can lead to surface spalls [223], [227].

Fracture and Cracking occur when the local stress concentration exceeds the tensile strength of the material of bearing components [223], [227]. Fracture is further classified in forced fracture and fatigue fracture. Forced fracture is primarily attributed to local overloading or over-stressing, while fatigue fracture is primarily attributed to repeated or cyclic bending often due to improper mounting [223], [227]. Generally, fracture often occurs in rings and cage of bearing elements. Cracking can also occur on bearing components as a result of excessive sliding or frictional heat [223], [227].

Although the exact percentages differ, Xu et al. [222] and one of the leading bearing manufacturers, SKF [228] identify rolling contact fatigue and wear as the most common failure modes of bearings, as can be deduced from the pie chart in Figure 5.2(a). Further significant failure modes are cracking and fracture, corrosion and plastic deformation, but accounting for less than 40% of the total represented failure modes. Electrical erosion is not represented in the pie chart because of its low probability of occurrence in most applications. Both frequently occurring failure modes are mostly caused by particle contamination and lubrication-related issues, such as insufficient lubrication, improper lubrication, incorrect lubricant, lubrication degradation [228], as depicted in Figure 5.2(b). Two further significant causes of bearing failures are improper use and incorrect assembly, accounting for 18% and 16%, respectively. Improper use of bearing includes wrong bearing selection and extreme operating conditions, while improper mounting, wrong fit and incorrect handling are represented under incorrect assembly [222], [228].

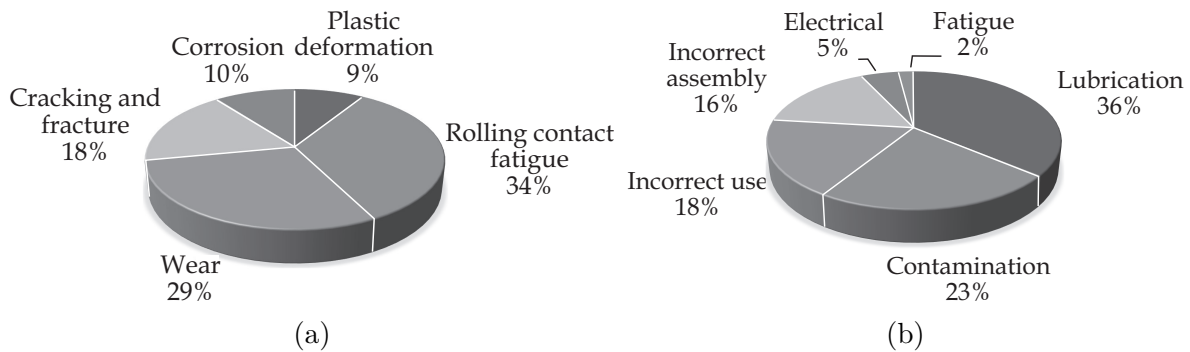


Figure 5.2: Statistics on (a) typical failure modes [222] and (b) corresponding causes [228]

5.1.2 Bearing monitoring techniques

After the identification of the failure modes, the next step that follows according to the ISO 17359 standard is the selection of appropriate measurement techniques, if the failure modes are measurable via physical parameters [26]. According to Moubray [229], different failure modes can show similar symptoms, such that the exact root cause of a failure is often exactly not identifiable. Under consideration of a vertical load, a steel ball bearing subjected to repeated stress can undergo different stages leading to fatigue failure, as exemplarily depicted in Figure 5.3.

The first stage is predominantly characterized by metal-to-metal contact during the bearing running-in period, where a new bearing stabilizes under the initially applied operating

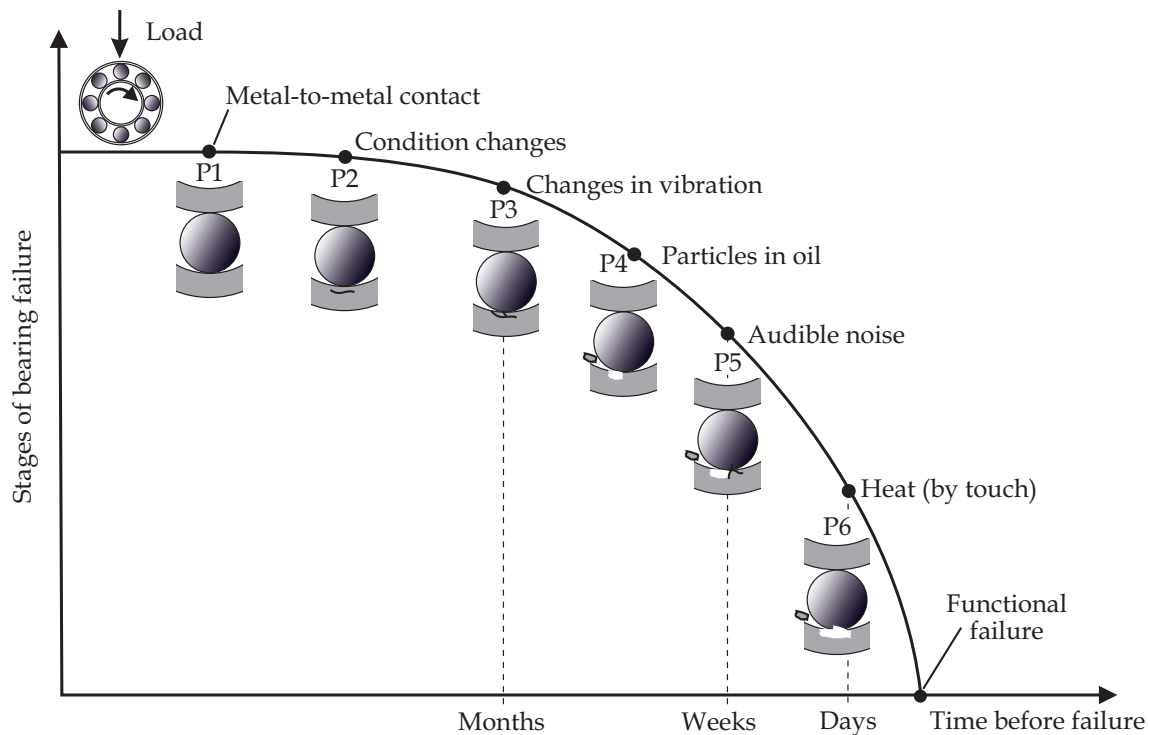


Figure 5.3: Development of bearing fatigue failure and some typical parameters, adapted from [229], [230], [231]

conditions [223]. In the second stage, during the bearing rotation, the outer/inner race moves microscopically and under cyclic stress, subsurface fatigue cracks initiate [229]. This subsurface cracks migrate to the surface of the race and form surface microcracks in the third stage producing slight changes in the vibration signature [229]. In the fourth stage, the microcracks develop into microspalls resulting from the over-rolling of the balls and peeling-off of metal particles from the raceways. The particles are detectable via oil analysis, while the microspalls are detectable by vibration analysis, during the over-rolling of the balls [229]. Stage five is characterized by spalling of the raceways and crack initiation on the balls. The spalls are large enough to emit an audible noise. A few days before the bearing fails, it ceases to run smooth and the temperature rises in the final failure stage [229].

Several measurement parameters are listed for different machine types and components in annex B of the ISO 17359 standard [26]. Typical parameters indicating bearing wear or damage are vibration, temperature and oil debris. Electric current is a further parameter specifically for electric motor faults [26], while acoustic emission is another commonly employed bearing monitoring technique also listed in annex D of the ISO 13379, part 1 standard [232]. In the following, the aforementioned measurement techniques are briefly described, with a focus on the ball bearing application example.

Vibration monitoring is one of the most frequently employed monitoring technique, because it provides a reliable and accurate assessment of the system condition [233] even at an early stage, as can be deduced from Figure 5.3. All system resonate at their resonance frequency and generate a certain vibration signature, even in normal state [234]. Deviation from the certain vibration signature in amplitude and/or frequency due to a fault can then be retraced to the specific system [234]. For bearings, amplitude changes

at the bearing characteristic frequencies. That is, the frequency related to a bearing component are often analyzed in the frequency domain to diagnose bearing faults [234]. Displacement, velocity and acceleration are the physical quantities of vibration typically measurable by different types of sensors for condition monitoring purpose. Further details on piezoelectric accelerometers, the other quantities and respective required transducers can be found in [234]. General information on vibration condition monitoring, processing and analysis can be found in parts 1 and 2 of the ISO 13373 standard [233], [235].

Acoustic emission is related to vibration monitoring, because it also considers waves emitting from a system [236]. However, as opposed to vibration, acoustic emission is related to transient elastic waves generated when a process, such as crack growth, releases spontaneous elastic energy [236]. Furthermore, the frequency range of acoustic emission is in the range [20, 1000] kHz, which allows an earlier and more reliable detection of faults [236], [237]. As opposed to vibration monitoring, which allows the detection of faults in stage three onward, that is, the detection of surface microcracks and pitting, acoustic emission typically allows the detection of developing faults from stage one onward, meaning it allows the detection of metal-to-metal contact to some extent, the initiation and further development of subsurface cracks [231]. As opposed to sound pressure technique that requires microphones to record sound waves in the frequency range up to 20 kHz [237], acoustic emission requires a specialized acoustic emission transducer or sensor. Although acoustic emission has numerous advantages as compared to vibration technique, the associated disadvantages, such as the high cost of required acquisition system and its vulnerability to attenuation and surrounding noise, hinders its widespread applicability [236], [237]. Further information on the application of acoustic emission for condition monitoring can be found in the ISO 22096 standard [236].

Lubricant, and mostly oil monitoring is mainly carried out for closed-loop lubrication systems, where the lubricant circulates in the system, such as in geared systems or non pre-greased bearings [234]. Lubricant monitoring allows the assessment of the state of the lubricant and the system's state [238]. To this end, six analysis are mainly conducted, namely viscosity analysis, oxidation analysis, water and acid content analysis, temperature analysis, particle counting and wear debris analysis [230]. These analysis are typically categorized under spectrographic oil analysis procedures, chip detectors, and ferrography. Spectrographic oil analysis procedures mostly encompasses the assessment of the lubricant's degradation state, for example via the kinematic viscosity, while chip detectors and ferrography typically utilize filters and magnetic plugs to separate wear debris or metal particles peeled-off from the raceways in the fourth stage onward [234]. Unlike vibration or acoustic emission, lubrication and wear debris analysis do not allow the isolation of the faulty component, because different components can be composed of the same material [239]. A further limitation is the high costs of the online monitoring equipment [230]. Most lubricated-for-life bearings are sealed, such that external particles or chemical contaminants hardly find their way into the bearing raceways [223]. Furthermore, the benefit-cost ratio is relatively low for such an expensive online monitoring of small bearings, such as utilized in this application example.

Temperature measurement considers the sudden temperature increase or temperature variation as an indicator of a developed fault [240]. The contact between the rolling element and raceway, viscous friction between inner ring and cage, friction between the rolling element and cage pocket and the drag force of the rolling element are the main sources of heat in rolling element bearings [241]. Since bearings are practically encased in a bearing housing system, the bearing temperature is often measured at the bearing outer race or

inferred from the bearing housing system, owing to the heat transfer process between the bearing and its housing [241], [242], [243]. Resistance temperature detector (RTD), thermistor and thermocouples are the most frequently employed contact sensors, while infrared thermography camera is the most commonly employed non-contact measurement technique [240], [244]. Generally, the bearing temperature is influenced by the viscosity of the lubricant, its load-carrying capacity, load distribution and power loss [243]. However, the bearing temperature is also influenced by the operating speed and load [238], [241], [243]. Further limitations of the temperature measurement are that it typically allows the detection of faults at the last stages of a bearings' life and the faulty component cannot be easily identified [234], [238].

Electrical measurement involves the acquisition of electrical parameter, such as current, voltage and torque [240]. Since electric motors also consists of bearings, then at this point it must be clearly defined that bearings encased within an electric motor are internal bearings and bearings not within the electric motor, such as those in a gearbox are external bearings [245]. Owing to the probability of failure occurrence and significant consequence of internal bearing faults, rotor cage winding faults and eccentric air gaps, the rotor and stator current are mostly monitored in induction electric motors and equipment driven thereof, such as gearboxes [246]. Motor drives emit a distinct current signature at normal state, such as in vibration. Thus, in motor current signature analysis, the deviation in frequency or amplitude of the monitored current signal are mostly analyzed in the frequency spectrum [245], [246]. Current monitoring and corresponding current signature analysis are mature techniques and have been successfully employed in diagnosing faults of internal bearings [247], [248], [249]. Several research have been conducted towards current monitoring and the application of current signature analysis in fault diagnostics of external bearings, because current monitoring is non-invasive, external sensors are not required and hence it is very cost effective [237], [245]. However, the signature of external bearing faults are often buried within the dominant supply frequency components, such that external bearing component faults are difficult to detect [237]. Furthermore, the signature of external faults can be masked as internal faults, which are often indistinguishable [246]. Although current monitoring and corresponding current signature analysis has found limited application in detecting external bearing faults due to the foregoing shortcomings, it is still under active research owing to the significant economical advantage.

RTD temperature contact sensors are used in this study to infer the bearing temperature, due to its low cost and availability. Moreover, stud-mounted piezoelectric accelerometer is also considered due to its wider dynamic and frequency ranges as compared to displacement and velocity transducers [234]. Considering that the ball bearing in this application example is grease-lubricated and lubricated-for-life, the lubricant and wear-debris monitoring is not taken further into account. The other described techniques are subsequently omitted, due to their shortcomings, such as high costs and infeasibility in the context of the considered application example.

5.1.3 Test rig development

A suitable dataset considering time-varying operating conditions, which emulates real-world situation is required to evaluate the proposed framework. In the light of promoting research, several open-source datasets have been published by different researchers in the PHM field. However, some consider a constant operating condition per experiment, such as pronostia [134], the bearing fault datasets [250] of the machinery failure prevention

technology (MFPT) and CWRU bearing datasets [251]. Although some datasets consider time-varying operating conditions, they possess several shortcomings. For example, Huang and Baddour [252] published a ball bearing dataset of different seeded faults, where vibration data was acquired under time-varying shaft speed. Although, they consider different profiles of time-varying speed per experiment, the dataset is applicable to evaluate only diagnostics methods, because the experiments are not run-to-failure, but only capture distinct health states. The same shortcoming applies to the Politecnico di Torino rolling element bearing dataset considering varying speed and load [253]. Klausen et al. [254] published a run-to-failure dataset of a ball bearing subjected to constant load and time-varying speed per experiment. Limitations of the dataset are that only an operating condition was considered and only one dataset was published, which limits the application of data-driven methods. A further dataset is a real-world dataset published by Del Campo Barraza [255], which considers six wind turbines subjected to time-varying speed profiles. Although the rotating speed vary within each wind turbine data, only one turbine had a bearing fault, one had an electrical fault, while others remained healthy. Due to the limited instances and because only one wind turbine had a relevant mechanical fault, an evaluation utilizing only the dataset might be bias. Furthermore, information pertaining to other operating conditions, such as the load or temperature, which influences the degradation of bearings was not provided. Thus, there is a need for a test rig to generate further run-to-failure datasets for evaluating diagnostics and prognostics algorithms under time-varying operating conditions without the aforementioned shortcomings.

The component under test and hence the main component of the test rig is a ball bearing, mounted on a shaft. The test rig is designed in accordance with the VDI 2221 Part 1 [256] under the minimum requirement that suitable actuators allow the operating conditions, namely the rotating speed and exerted load to be automatically varied during an experiment, while acquiring condition monitoring signals with respective sensors. The function structure diagram representing the inputs and outputs of a system, as described in Pahl et al. [257] and Bender et al. [258], is depicted in Figure 5.4 for the bearing test rig. The main functions are exerting a static (F_{stat}) and dynamic (F_{dyn}) load on the bearing, driving the bearing, monitoring the condition of the bearing and absorbing the reaction forces and moments. The inputs are the power supply to the actuators and sensors, the set values of the static load, the dynamic load and the rotating speed.

Apart from the dissipated energy, measured system outputs are the output static load, output dynamic load, output rotating speed, the bearing temperature and the vibration of the bearing. The main functions of the bearing test rig are elaborated upon, in the following.

To ensure adequate support of the shaft, two additional bearings for a locating (fixed)/non-locating (free) bearing arrangement must be integrated into the design [259]. In the following, the term *support bearing* is introduced for the locating/non-locating bearing and the term *test bearing* corresponds to the bearing under test, to avoid ambiguities. To obtain sufficient run-to-failure data for validating the proposed methodology, the loads acting on the test bearing are intended to accelerate its degradation. However, to avoid impairment and falsification of the test bearing monitoring, it is important to also ensure that the support bearings are designed to be as fatigue-resistant as possible so that no signs of degradation occur during the run-to-failure experiments of the test bearing. Thus, a precise selection of the test bearing can only be made in combination with the selection of the support bearing. To ensure a uniform load distribution and consequently symmetrical deflection, the support bearings should be on both sides of the test bearing

and the design of the shaft between the support bearings should also be symmetrical. Thus, the two support bearings are arranged at an identical distance from the test bearing, as illustrated in Figure 5.5, such that the force is evenly distributed on each of the two support bearings. This symmetrical requirement implies that the geometric dimensions of the support bearings must be identical or be of the same type.

Although spherical roller bearings and self-aligning ball bearings can compensate for unavoidable shaft misalignment up to approximately 4° , spherical roller bearings can support higher radial and axial forces and are therefore more resilient. Thus, they are considered as support bearings [260]. Deep groove ball bearings with suitable dimensions are considered as a test bearing, because they are most cost-effective rolling element bearings with relatively low load ratings and the highest availability [259]. As the load carrying capacity is roughly proportional to the size of a bearing, the dimensions of the test bearing should be as small as possible in order to keep the required operating loads low and

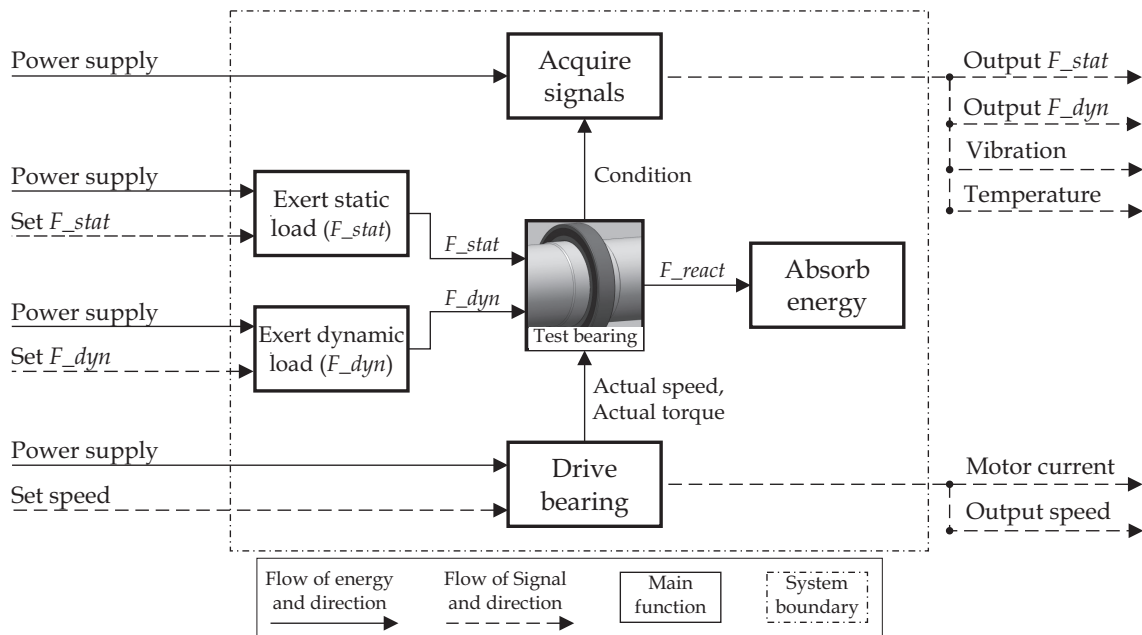


Figure 5.4: Function structure diagram of the bearing test rig adapted from [259]

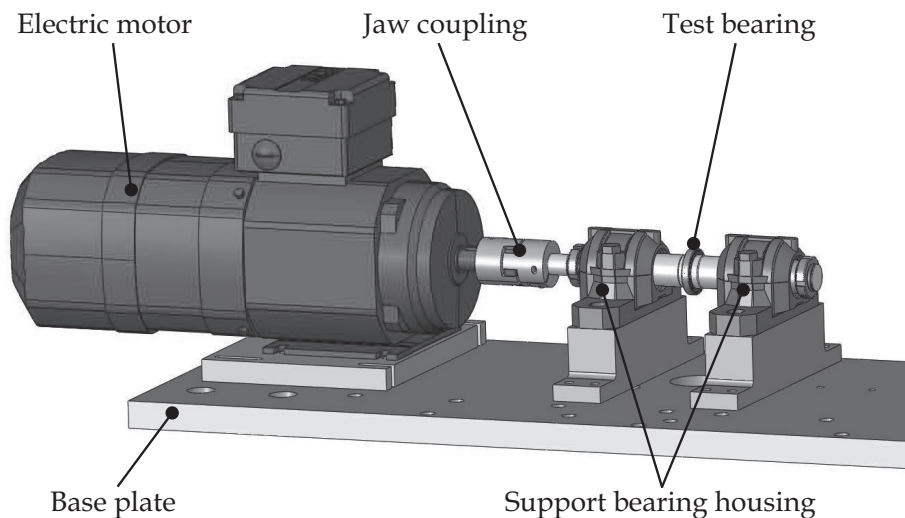


Figure 5.5: Main bearing test rig components for driving bearing

thus reduce operating costs [259]. The test bearing housing, which encases the bearing is designed as a split housing to ensure quick replacement. At the same time, the bearing housing serves as a connecting element for the acting loads, such that the applied forces are transferred to the test bearing through the bearing housing. Furthermore, because of the direct inaccessibility of the test bearing, the bearing housing serves as an interface in monitoring the condition-related signals, such as vibration and temperature. During operation, vibration signals of the test bearing are transmitted directly to the sensors through the bearing housing and the bearing temperature can be determined directly in the vicinity of the test bearing [259]. The test bearing is a deep groove ball bearing of type 61806-2RS with a bore diameter of 30 mm, an outer diameter 42 mm, a width of 7 mm, a basic static load rating C_0 of 3.15 kN and a dynamic load rating C of 4.0 kN, as also presented in Table 3 in the Appendix. The support bearings are spherical roller bearings of type WS22205-E1-XL-2RSR with a bore diameter of 25 mm, an outer diameter of 52 mm, a bearing width of 23 mm, a basic static load rating C_0 of 42.5 kN and a dynamic load rating C of 48.5 kN. Under consideration of the resulting dimension and material of the shaft, the safe working load of the shaft is 5 kN with a fatigue safety factor of 2.13 [259].

The function block *acquire signals* defines an interface that bundles the measured system signals in form of the operating conditions and condition-related monitored signals. As presented in Figure 5.4, although the system operating conditions set values are the inputs to the system, their actual values are measured by appropriate sensors for control purpose and for further data processing steps, such as required when considering time-varying operating conditions. The condition monitoring signals pertaining to the health of the test bearing are the vibration, temperature and motor current. The measured system signals are elaborated in the following section.

The function block *drive bearing* provides the basic functionality of driving the bearing, which sits on a shaft. To this end, an asynchronous motor is connected to the shaft via a jaw coupling, as illustrated in Figure 5.5. Electric motors basically consists of a stator and a rotor. The stator is stationary, while the rotor rotates in the electric motor as a result of the current flowing through it. The rotor of an asynchronous or induction motor is typically a short-circuit squirrel cage and the stator is a three-phase coil winding [261]. The rotor rotates when an AC passes through the coil on the stator. During operation, the asynchronous motor is typically connected to the three-phase AC source and thus generates a magnetic rotating field in the coil winding of the stator. The rotation of the magnetic field consequently rotates the rotor. The motor motion depends on the difference in the rotational speed of the magnetic field in the stator and the rotational speed of the rotor, that is, the slip [261]. A positive slip, namely, an asynchronous speed between the rotor speed and the rotational speed of the magnetic field is required to generate torque and hence for the motor to run. The speed of the asynchronous motor depends on the frequency of the rotating field [259], [261]. Therefore, frequency converters are required to control the motor, because they convert a sinusoidal AC supply voltage into an AC voltage with variable frequency and amplitude. Furthermore, frequency converters are usually equipped with integrated measurement techniques to measure the resulting motor current without additional sensors [259], [261]. The asynchronous AC motor of type DRL80S4 from SEW-EURODRIVE with a speed of 3000 rpm and a torque of 6 N m is utilized in the test rig.

The function block *absorb energy* compensates for the resulting forces and moments of the test rig. Furthermore, it promotes the stability of the entire test rig and includes a safety

mechanism to protect the user [259]. This is achieved via a massive base plate, where the other components of the test rig are mounted, under consideration of the geometry of the other components. Furthermore, shields are mounted on the base plate to prevent possible loose parts of rotating components from causing damage to the user or other components during an experiment. Apart from providing a safety mechanism, the shield also allows the visual inspection of critical test rig components during experiments [259].

The function block *exert dynamic load* (F_{dyn}) provides the test rig with the functionality of exerting time-varying radial load on the bearing under test. Although several principles, such as hydraulic, pneumatic and electrodynamic drives are conceivable to fulfill the required functionality, an electrodynamic shaker was selected [259]. The working principle of an electrodynamic shaker is similar to that of a loudspeaker [262], [263]. A coil of wire, also known as a voice coil, is surrounded by a constant radially acting magnetic field within the shaker [262], [263]. When current is passed through the voice coil, an axial force proportional to the current is generated, which consequently moves the table structure fixed to the shaker [262], [263]. By reversing the current direction, the direction of the generated force is also reversed. With a given orthogonality of the current, magnetic field and force vectors, the Lorentz force generated is proportional to the current strength, the magnetic flux density and the length of the coil of wire [262], [263], [264]. The connection between the shaker and the test bearing housing is realized via a stinger. The shaker is directly connected to the stinger holder, while there is a force sensor between the stinger and the bottom test bearing housing to directly measure the applied force, as shown in Figure 5.6(a). Specifically an electrodynamic shaker from Spectral Dynamics of type SD-2420 is utilized in the test rig. The electrodynamic shaker is specified for a sine peak force of approximately 10 kN, a usable frequency of approximately 3 kHz and a maximum peak-to-peak displacement of 51 mm.

The function block *exert static load* (F_{stat}) provides the test rig with the functionality of exerting a static pre-load on the bearing under test, while guaranteeing support during dynamic loading with the electrodynamic shaker. Linear actuators convert rotation into translation motion for the purpose of displacement or force exertion [265]. Depending on the type of medium utilized for conveying energy to the actuating system, linear actuators can generally be categorized in hydraulic, pneumatic and electrical actuators [265].

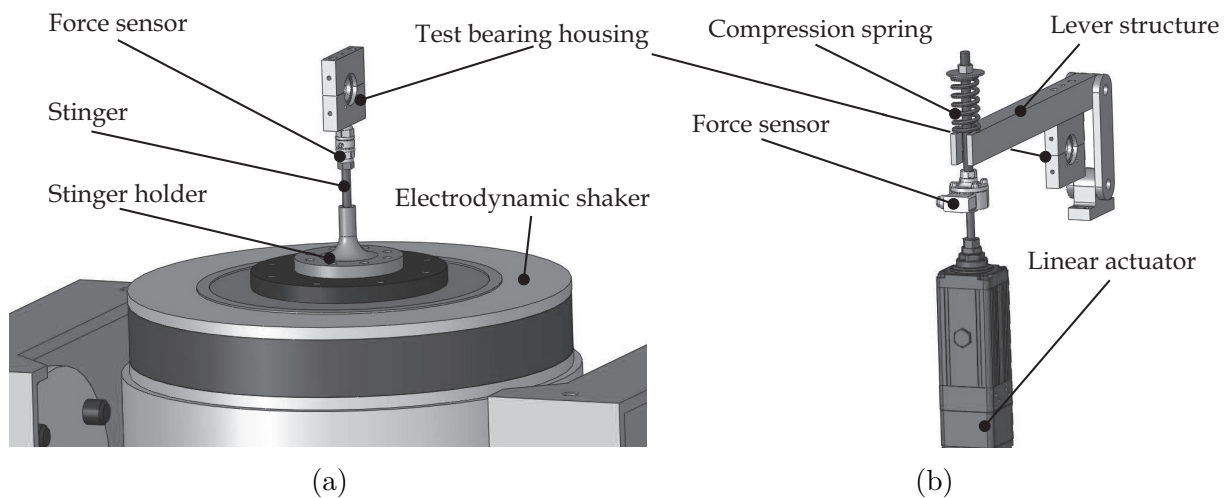


Figure 5.6: Main bearing test rig components for exerting (a) dynamic load and (b) static pre-load

Although hydraulic actuators are mostly used in industry because it delivers high force at a relatively low cost than electrical actuators, it is more bulky, noisy, requires regular maintenance and its fluid can pollute the environment [266], [267]. Advantages of an electrical actuator are its flexibility in control of position, velocity, acceleration or deceleration and output force. Furthermore, it only uses as much power as needed at any given time, thus it is power-efficient [266], [267]. Owing to the aforementioned advantages and due to the compact design of an electric linear actuator, which allows an uncomplicated integration into the test rig, a rod style electric linear actuator is further considered. The linear displacement or stroke of the rod is intended to compress a compression spring on the path to the bearing, thereby generating a spring force, whose magnitude depends on the stiffness of the compression spring [259]. The compression spring is mounted on one end of a lever structure, while the test bearing housing is attached to the other end of the lever structure, such that the resulting force is exerted on the bearing via a lever structure, as illustrated in Figure 5.6(b). A CMSMB50 ball screw rod-style electric linear actuator from SEW-EURODRIVE with a peak feed force of 8 kN and a stroke length of 70 mm is utilized in the test rig. In summary, the assembled bearing test rig with its components is depicted in Figure 5.7(a).

As discussed, the shaft is coupled with the driving motor (1) via a jaw coupling (2) and supported by two spherical roller bearings (8) within their housing. Furthermore, the test bearing is mounted on a shaft and encased within its housing (3), where accelerometers (4) and Pt100 sensors are mounted. A static pre-load is exerted on the bearing via a lever structure (5), which is attached to the upper part of the bearing housing. To this end,

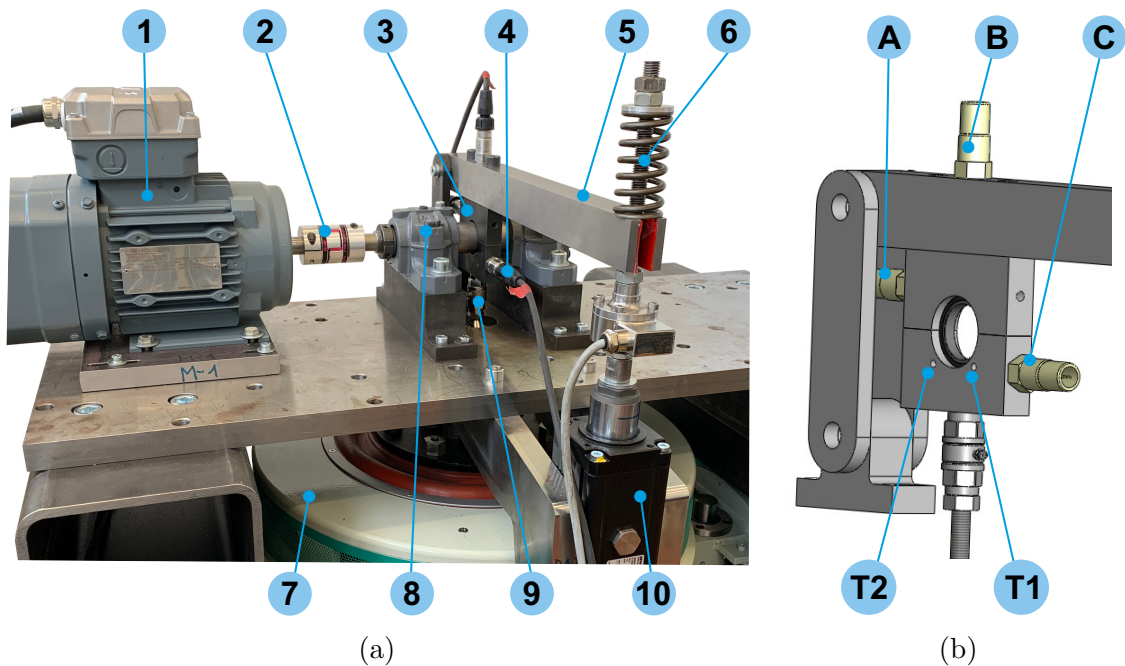


Figure 5.7: (a) Assembled ball bearing test rig with the following components: (1) electric motor, (2) jaw coupling, (3) test bearing housing, (4) accelerometer, (5) lever structure, (6) compression spring, (7) electrodynamic shaker, (8) support bearing housing, (9) stinger connected to a quartz force sensor, and (10) linear actuator.

(b) Accelerometer placement and temperature position on the test bearing housing without the shaft, adapted from [169]

the compression spring (6), mounted on the lever structure, is compressed by the linear actuator (10). A dynamic load is superimposed on the static pre-load by means of an electrodynamic shaker (7), which is connected to the test bearing housing via a stinger (9) [169].

5.1.4 Measurement setup

The measurement setup for the test rig can broadly be categorized in actuator and sensor measurement setup. Generally, the measurement setup comprises of a data acquisition (DAQ)-software, a DAQ-hardware, a signal converter, a signal conditioner and an actuator or sensor. The DAQ-software allows the processing, such as a control program or the storage of data. MATLAB® is used as the programming language with its corresponding DAQ-software package. Although data and signal are used interchangeably, data is information collected, processed and stored by a computer, while signal is information transferred by a physical medium, such as a copper wire [268]. The DAQ-hardware implies the hardware that directly interfaces with the DAQ-software via a dedicated communication channel, here. A National Instruments (NI) CompactDAQ cDAQ-9189 ethernet chassis is utilized here as the DAQ-hardware. It controls the timing, synchronization, and data transfer between specific input or output modules and the DAQ-software over the ethernet network [269]. The signal converter can be a digital-to-analog converter (DAC) or a analog-to-digital converter (ADC), depending on if the input signal or required output signal is digital or analog, respectively. The utilized signal converters are listed in the subsequent paragraphs. Signal conditioning includes external voltage or current excitation, amplification and filtering [270]. The actuators are the electric motor, the electrodynamic shaker and the linear actuator, while the sensors that measure the actual speed, the actual forces, the vibration, the temperature and motor current are elaborated on, in the following.

A general measurement chain for the employed actuators, as depicted in Figure 5.8 are briefly described, in the following.

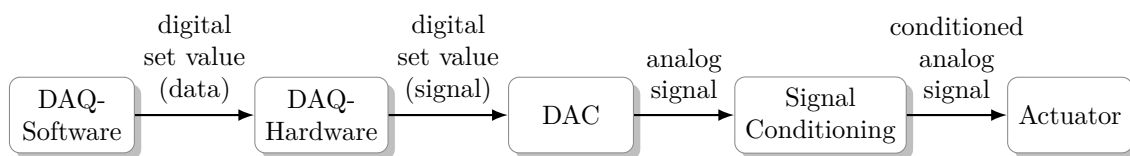


Figure 5.8: Typical measurement chain for actuators

Under consideration of the speed limits of the components of the test rig, the set value of the rotating speed as well as the rotating direction of the electric motor are defined in the DAQ-software, MATLAB® and passed on to the DAQ-hardware, cDAQ-9189. The set rotating speed signal is then transmitted to a NI-9263 DAC module, which converts the input signal to a voltage output, while the rotating direction signal is transmitted to a NI-9472 digital output module. The resulting output signals are consequently transmitted to the DIO11B input/output board of the frequency converter and then to the electric motor, which rotates ideally at the specified speed and direction. The measurement chain for the linear actuator is similar to that of the electric motor. However, the major difference is that the inputs are a set static force alongside the up/down direction, as opposed to the set speed and left/right direction of the electric motor. Under consideration of the static force, the bearing load and the load limits of other test rig components, the set dynamic

force is defined in MATLAB[®] and passed on to the cDAQ-9189. The set dynamic force signal is consequently transmitted to a NI-9263 module, which converts the input signal to a voltage output. The resulting output signal is then transmitted to the DA-10 power amplifier from Spectral Dynamics, which amplifies the output signal before it is finally been passed on to the electrodynamic shaker.

Similar to the measurement chain for the actuators, the measurement chain for the sensors is depicted in Figure 5.9 and is described in the following.



Figure 5.9: Typical measurement chain for sensors

The electric motor has an internal encoder of type ES7S, which measures the output speed of the motor. The sensed signal is transmitted to the frequency converter for signal conditioning and then transmitted consequently to the ADC NI-9221 analog input module, which converts the analog input signal to a digital signal. The digital voltage signal is transmitted to the cDAQ-9189 DAQ-hardware and finally transferred to MATLAB[®], where the signal is converted to speed. The output static load is measured by a U2B force sensor from Hottinger Brüel & Kjær GmbH (HBK), formerly Hottinger Baldwin Messtechnik GmbH (HBM), with a nominal (rated) force of 2 kN. The sensed signal is amplified by an HBM MC55 and then transmitted to the ADC NI-9221 analog input module. Consequently, the digital voltage signal is transmitted to the cDAQ-9189 DAQ-hardware and then to the MATLAB[®] DAQ-software. The output dynamic load is measured by a quartz force link sensor of type 9321B from Kistler, with a force range of ± 10 kN and amplified by a respective Kistler amplifier of type 5011. The subsequent steps in the measurement chain, namely, the NI-9221, the cDAQ-9189 and MATLAB[®] is similar to the measurement of the output static load. The three phase, namely the *U*-, *V*- and *W*-phase currents between the frequency converter and the electric motor are measured by current transducers of type LTS 25-NP from LEM with a primary nominal RMS current of ± 25 A and a nominal output voltage of 2.5 ± 0.625 V. The sensed analog signal is digitized by the ADC NI-9221 and consequently transmitted to the cDAQ-9189 DAQ-hardware and respective DAQ-software.

The test bearing temperature is inferred from the bearing housing temperature. To this end, two 3-wire, 100 Ω RTD sensors, also known as Pt100 sensors are placed in through bores at the bottom part of the bearing housing in close proximity to the outer ring of the test bearing, as depicted in Figure 5.7(b). Furthermore, the ambient temperature is also measured with two additional Pt100 sensors. The measured analog temperature signal is transmitted to the NI-9216 temperature input module, where the analog signal is conditioned, such as excited with an electric current and amplified. Furthermore, the conditioned signal is also converted to a digital signal with the integrated ADC within the NI-9216 temperature input module. The digitized signal is consequently transmitted to the cDAQ-9189 DAQ-hardware and then to the MATLAB[®] DAQ-software at a sampling frequency of 10 Hz. The Pt100 sensors are of type MC 745691-02 from Measurement Computing with a temperature measurement range of $[-50, 204]$ °C. Three single axis accelerometers are stud-mounted on the test rig for vibration monitoring of the test bearing. One is mounted on lever structure to measure the vibration vertically,

one is mounted horizontally on the upper part of the bearing housing, the other is also horizontally mounted, but on the bottom part of the bearing housing, as shown in Figure 5.7(b). The accelerometer placement in the horizontal and vertical direction is as suggested by the ISO 13373 standard [233]. The necessary current source to power the integrated circuit piezoelectric (ICP[®]) accelerometers is provided with the Kistler power supply/coupler of type 5134B. The conditioned vibration signal is digitized with the NI-9223 input module and consequently transmitted to the cDAQ-9189 DAQ-hardware and then to the MATLAB[®] DAQ-software. To facilitate high frequency analysis, the vibration signals are sampled at a frequency of 128 kHz for some experiments and due to data storage issues at 64 kHz for other experiments. The accelerometers are low-cost industrial ICP[®] accelerometers from PCB[®] Piezotronics of model type 603C01 with a frequency range of [0.5, 10 000] Hz and a measurement range of ± 50 g.

During experiments, a measurement is acquired at a sampling duration of 1.6 s and at an interval of approximately 12 s between measurements. With the exception of the temperature and vibration signals, the other sensor signals are connected to the same NI-9221 module. Hence, the sensor signals are sampled at the same frequency of 10 kHz.

5.1.5 General experimental procedure

The goal of the following experiment is to generate run-to-failure condition monitoring data of ball bearings subjected to time-varying speed and load profiles for the validation of the proposed methodology. The basic idea of implementing time-varying operating condition during the run-to-failure experiments is to set a random speed and load amplitude from a predefined distribution under consideration of predefined speed and load limits, similar to the profiles described by Villa et al. [271]. During the experiment planning phase, the popular pronostia dataset is taken into consideration because it also considers run-to-failure experiments of deep groove ball bearings, but of type NSK 6804DD with a basic dynamic load rating C of 4 kN and a limiting speed of 13 000 rpm [134], [272]. To achieve an accelerated degradation of the ball bearings, the maximum load considered in the pronostia dataset is 5 kN, which results in a C/P , that is, a ratio of dynamic load rating to equivalent dynamic load capacity of 0.8 and an overload of 1 kN. Furthermore, the pronostia dataset considers about 12% of the limiting speed as the minimum speed. Thus, owing to the safe working load of the shaft and similar bearing characteristics of this application example to the pronostia test bearing, the maximum load is limited to 5 kN during the experiments. Furthermore, to achieve a reasonable duration of a run-to-failure experiment and consequently achieve a minimum cost of conducting the experiments, the minimum load is set to 3 kN. During the experiments, the maximum speed is limited by the speed of the support bearings and the nominal speed of the motor to 3000 rpm and the minimum speed is set to 1500 rpm, which is also approximately 12% of the limiting speed of the test bearing. Specifically, the speed is relatively constant during a measurement k , as defined in equation 5.2.

$$n_k(t_l) = n_{set,k} + \varepsilon_{speed,k}(t_l), \quad (5.2)$$

where n_k is the output speed at the k^{th} measurement, sampled with a sampling frequency of $f_{s_speed,k}$ and measured with a total measurement duration of $M_{dur,k}$. The l^{th} time sample is denoted by t_l , $l \in [1; S_k]$, $S_k := f_{s_speed,k} \cdot M_{dur,k}$ is the total number of samples, and $\varepsilon_{speed,k}$ is a random measurement noise. However, the set speed $n_{set,k}$ is randomly sampled

from a stationary uniform distribution within the interval [1500, 3000] rpm between measurement intervals. Additionally, the total load on the bearing during a measurement k is a superimposition of a relatively constant static pre-load $F_{stat,k}$ and a dynamic load $F_{dyn,k}$ component. Random and sinusoidal dynamic loads are considered as exemplarily shown in Figure 5.10, to meet random or cyclic real-world loading situations.

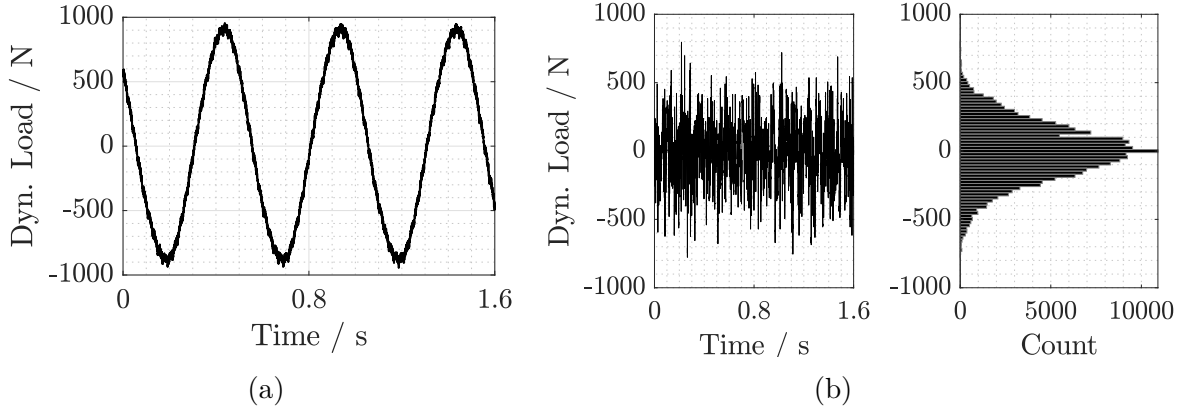


Figure 5.10: Representative dynamic load exerted during the experiments: (a) sinusoidal load and (b) Gaussian load, adapted from [169].

Between measurement intervals, the set static pre-load $F_{stat_set,k}$ and the set dynamic load amplitude $F_{amp_set,k}$ are also randomly sampled from stationary uniform distributions within intervals [1500, 4500] N and [50, 1500] N, respectively. Furthermore, the total load amplitude on the bearing, that is, $F_{stat_set,k} + F_{amp_set,k}$ lies within the defined load limits of [3, 5] kN. The static pre-load at the k^{th} measurement is defined in equation 5.3.

$$F_{stat,k}(t_l) = F_{stat_set,k} + \varepsilon_{stat,k}(t_l), \quad (5.3)$$

where the output static pre-load $F_{stat,k}$ is sampled with a sampling frequency of $f_{s_stat,k}$ and measured with a total measurement duration of $M_{dur,k}$. t_l denotes the l^{th} time sample, and $\varepsilon_{stat,k}$ is a random measurement noise. The sinusoidal dynamic load at the k^{th} measurement is defined in equation 5.4.

$$F_{dyn,k}(t_l) = F_{amp_set,k} \sin(2\pi f_{dyn} t_l) + \varepsilon_{dyn,k}(t_l), \quad (5.4)$$

where the output sinusoidal dynamic load $F_{dyn,k}$ is sampled with a sampling frequency of $f_{s_dyn,k}$ and measured with a total measurement duration of $M_{dur,k}$. The constant frequency of the sinusoidal load is $f_{dyn} := 2 \text{ Hz}$ and $\varepsilon_{dyn,k}$ is a random measurement noise. A random dynamic load cannot be easily described as a function of time, because the value of the dynamic load $F_{dyn,k}(t_l)$ is not predetermined at a given time t_l [4]. However, the random dynamic load can be described via its probability density function (PDF). The random dynamic load is Gaussian, truncated at the lower and upper truncation points, $y_{lt} := -F_{amp_set,k}$ and $y_{ut} := +F_{amp_set,k}$, respectively. The PDF of a truncated Gaussian random variable Y_k with mean μ and variance σ^2 is expressed as in equation 5.5 [273], [274]:

$$f_{\tilde{Y}_k}(y_k) = \begin{cases} \frac{f_{Y_k}(y_k)}{\int_{y_{lt}}^{y_{ut}} f_{Y_k}(y_k) dy} & y_{lt} < y_k < y_{ut} \\ 0 & \text{otherwise} \end{cases}, \quad (5.5)$$

where

$$f_{Y_k}(y_k) = -\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2}\left(\frac{y_k-\mu}{\sigma}\right)^2}. \quad (5.6)$$

A run-to-failure experiment is conducted under the described time-varying operating conditions until a predetermined failure threshold is exceeded. The failure threshold is defined under consideration of two aspects. Firstly, such that a permissible fault level of the test bearing is not exceeded under laboratory conditions and lastly, such that the failure of other components of the test rig, except the test bearing, is avoided [169]. The failure threshold is an inclusive disjunction of a vibration-based and temperature-based threshold criteria. The threshold recommendations in the withdrawn DIN ISO 10816-3 standard [275] is not directly applicable because of its validity for machines fault types in the frequency range of [10, 1000] Hz under stationary conditions [275], [276]. Furthermore, it builds on vibration displacement or velocity measurements and not on vibration acceleration, as measurable in the experiments. The stopping criteria for the pronostia [134] and XJTU-SY [277] bearing run-to-failure experiments is set to 20 g and it is solely based on the vibration acceleration data for both experimental datasets. Such a constant vibration acceleration threshold value is not feasible over all experiments, because the operating conditions vary over time in this application example. Thus, a factor-based vibration acceleration threshold, considering the steady state of an healthy bearing is employed as the vibration-based criteria. It builds on the equivalent energy content of vibration acceleration measurements and is formulated as [169]:

$$A_{thresh} = O \cdot \frac{1}{m} \sum_{k=1}^m \text{RMS}(a_k(t_1), \dots, a_k(t_{S_{V,k}})), \quad (5.7)$$

where A_{thresh} denotes the vibration-based threshold value, $O = 8$ is a predetermined constant factor obtained via a systematic study to avoid failure of neighboring components, m is the number of vibration measurements considered in the steady healthy state, and $a_k(t_i)$ represents the i^{th} time index of a vibration signal sampled with a sampling frequency of $f_{s_a,k}$, measured with a total measurement duration of $M_{dur,k}$ and with a total number of $S_{a,k} := f_{s_a,k} \cdot M_{dur,k}$ samples. The vibration-based threshold value lies approximately between 3 g and 10 g in the experiments. The second failure criterion of the test bearing builds on the bearing temperature, because insufficient lubrication of the bearing raceway leads to increased friction and then to increased temperature of the bearing [169]. Furthermore, the bearing temperature also rapidly increases during the last stages of bearing failure. Thus, to additionally monitor the degradation of the test bearing and to avoid melting of the nitrile butadiene rubber (NBR) seal of the test bearing, the temperature-based threshold value T_{thresh} is set as 110 °C under consideration of the temperature measured at the bearing housing [169].

5.2 Data understanding and preparation

Three sets of experiments are considered to investigate the effects of the operating conditions. The first set of experiments consider the influence of the operating conditions on the condition monitoring signals. The second set of experiments consider the influence of the operating conditions on the failure mode. Lastly, the influence of the operating conditions on the degradation curve is also considered. The set of experiments are described in the following subsections.

5.2.1 Influence of the operating conditions on the condition monitoring signals

Two preliminary run-to-failure experiments were conducted to investigate the influence of the operating conditions on the condition monitoring signals. In the first experiment BS01, the rotating speed varies over time, while the amplitudes of the static and dynamic loads are constant over the run-to-failure experiment. In the second experiment BS02, the rotating speed and dynamic load amplitude are constant, while the static pre-load varies over time. Furthermore, in the following, the acquired data from both experiments are analyzed in the time-domain.

Preliminary experiment BS01: varying speed, constant static pre-load and constant dynamic load amplitude

The set rotating speed $n_{set,k}$ of the single run-to-failure experiment BS01 is uniformly sampled from six distinct values, such that $n_k \in \{200 \cdot i : i \in \mathbb{N}, 7 < i < 14\}$ rpm. Each measured distinct output speed is depicted in Figure 5.11(a) in different colors to easily identify the six distinct speed values. As presented in Figure 5.11(a), the measured mean output speed is approximately the set speed values for experiment BS01. The set static pre-load is constant at approximately 3000 N, while the set dynamic load is sinusoidal with a constant amplitude of approximately 1000 N and a constant frequency of 2 Hz over the run-to-failure experiment. As shown in Figure 5.11(b), the measured mean static load increases slightly during the bearing run-in period and approximately takes the set value afterwards. As opposed to the measured mean static load, the measured peak dynamic load is noisy, but also approximately assumes the set dynamic load value, as illustrated in Figure 5.11(c). The corresponding measured condition monitoring signals are depicted in Figures 5.12 and 5.13.

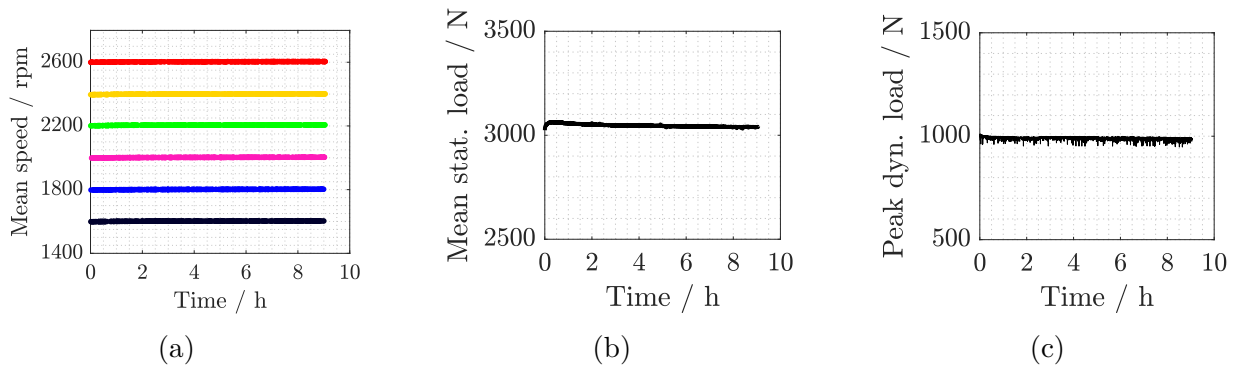


Figure 5.11: Measured operating conditions of the experiment BS01: (a) varying speed, (b) constant static pre-load and (c) constant dynamic load amplitude

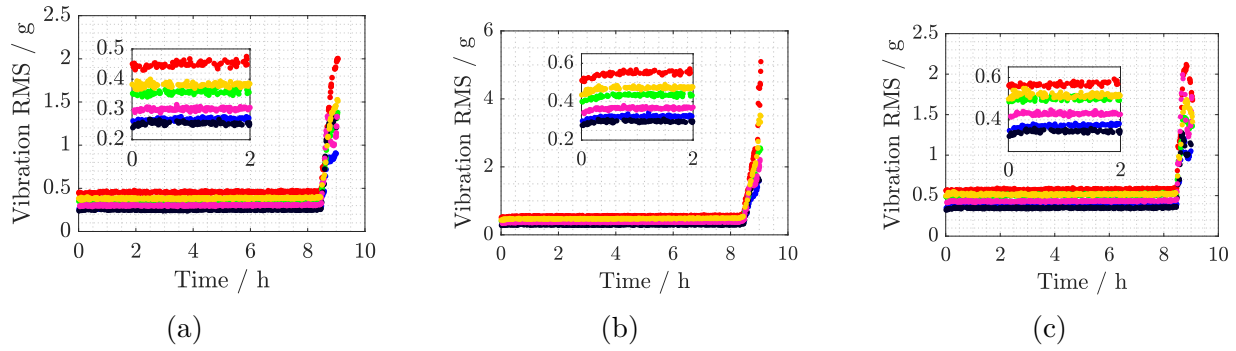


Figure 5.12: Vibration measurements of the run-to-failure experiment BS01: (a) horizontal accelerometer *A*, (b) horizontal accelerometer *C*, and the (c) vertical accelerometer *B*. The different colors correlate to the six distinct speed levels

The RMS of the vibration signals measured with the horizontal accelerometers *A* and *C* are shown respectively in Figures 5.12(a) and 5.12(b), while the vertical accelerometer *B* is shown in Figure 5.12(c). As can be seen in the figures, the variation in speed is reflected in the RMS of the measured vibration data, where the six different colors of the distinct speed values are mapped to the vibration data. As also can be seen in the figures, the vibration RMS is almost constant in each speed regime till about 8.5 h. Thereafter, the vibration RMS then abruptly increases till the end of experiment, where the predetermined vibration threshold value A_{thresh} is exceeded. The increase in the vibration RMS value with speed supports the findings of Williams et al. [278]. Although the ground truth degradation state is not directly measured, it can be inferred from the monitored signals, where the almost constant phase up to approximately 8.5 h can be considered an healthy state followed by a faulty state till the bearing EoL is reached. The degradation curve with an almost constant phase and a subsequent abrupt increase in the measured vibration signal is common to accelerated life tests of rolling element bearings, because of the increased contact stress in the rolling element contact zone due to the extreme operating conditions [134], [278], [279], [280], [281]. Furthermore, in the work of Kotzalas and Harris [282], spall progression on rolling element bearings was related to the vibration acceleration acquired with accelerometers, where two regions of the degradation curve were also evident. The RMS value of the vibration signal measured with the horizontal accelerometer *C* significantly increases in the faulty state monotonically, which represents the progression of bearing fault in the different speed levels. Conversely, in the faulty state, the RMS values does not monotonically increase in some speed regimes for the other horizontal accelerometer *A* and in all speed levels for the vertical accelerometer *B*. Furthermore, the vertical accelerometer *B* is placed on the lever structure, where the static force is exerted and hence is farther away from the test bearing and more susceptible to signal interference, due to the longer signal transmission path. Thus, the vibration signal measured with the vertical accelerometer *B* is not further considered in the following subsections.

The bearing temperature is inferred from the bearing housing temperature measured at locations *T1* and *T2*. Both measured absolute temperatures and corresponding relative temperatures under consideration of the ambient temperature are depicted in Figures 5.13(a) and 5.13(b), respectively for the run-to-failure experiment. As can be seen in both figures, the temperature trend and values are similar at both sensor locations, with a mean temperature difference of approximately 0.5°C for the run-to-failure experiment.

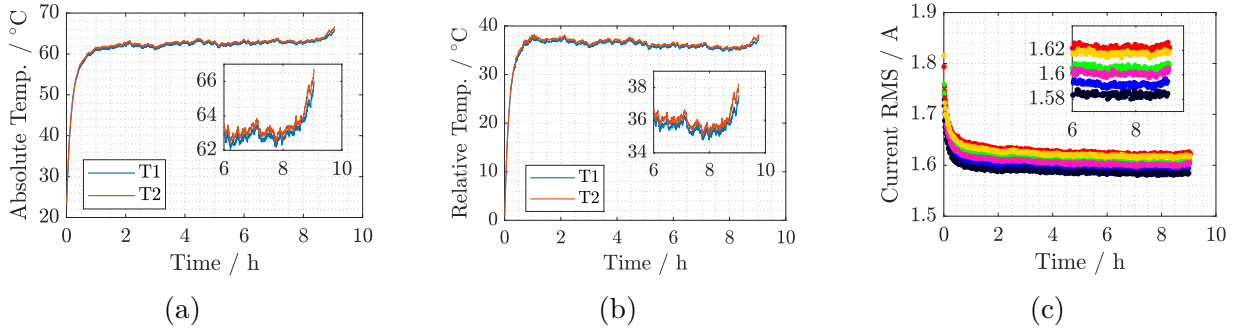


Figure 5.13: Temperature and motor current measurements of the run-to-failure experiment BS01: (a) absolute temperatures (b) and corresponding relative temperatures at sensor positions $T1$ and $T2$. (c) Motor current, where the different colors correlate to the six distinct speed levels.

In the following, only the temperature measured at the position $T2$ is further discussed, because the mean temperature difference at both locations is negligible. As can be inferred from Figure 5.13(a), the temperature curve can be partitioned in three distinct regions [283]. The first region is characterized by an initial temperature rise, where the absolute temperature $T2$ rises from the ambient temperature of approximately 23°C to the bearing operating temperature of approximately 63°C , from the onset of the experiment. In the second region, also known as saturation or steady-state region [238], [284], the temperature fluctuates slightly around the bearing operating temperature over a long period of time. In the third region, which is typically characterized by a progressive increase in temperature [283], the measured temperature slightly increases from 63°C to approximately 67°C at the end of the experiment. As can be deduced from Figure 5.13(b), the slight increase towards the end of experiment is attributed to the bearing temperature due to a fault and not due to the ambient temperature. In the findings of Dong et al. [285], the bearing temperature increased by approximately 2°C for every 2000 rpm increment in rotating speed, applied in four steps from 2000 rpm to 8000 rpm. Contrarily, the temperature fluctuation is less than 1°C in the considered run-to-failure, with a maximum speed variation of 1000 rpm.

The raw signal of the measured U -, V -, and W -phase motor currents from the same supply source are sinusoidal with a phase of 120° apart [286], [287]. Thus, only the U -phase is further considered here. As can be inferred from Figure 5.13(c), the RMS of the motor current can also be partitioned in two regions as in the temperature signal. However, in the first region, the RMS of the motor current decreases in the bearing wear-in period and stabilizes thereafter. Similar to the vibration signal, the RMS of the motor current increases with speed. However, different from the temperature and vibration signals, the RMS current value does not show any significant change pertaining to fault development or failure of the bearing, even at the end of experiment.

Preliminary experiment BS02: constant speed, varying static pre-load and constant dynamic load amplitude

The second preliminary experiment involves a run-to-failure experiment with a constant rotating speed of 2700 rpm, a varying static pre-load between 1500 N and 4500 N and a random dynamic load with a constant peak of 1500 N. Figure 5.14(a) indicates that the mean speed is approximately the set value of 2700 rpm over the entire duration of the experiment. The mean static pre-load varies randomly between [1500, 2500] N from the

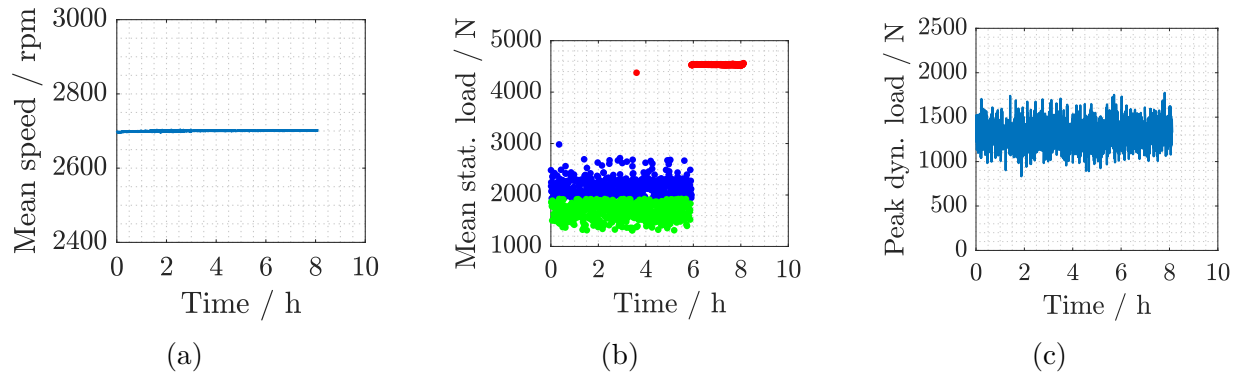


Figure 5.14: Operating conditions of the run-to-failure experiment BS02: (a) constant speed, (b) varying static pre-load and (c) constant random dynamic load amplitude

beginning of experiment till approximately the sixth hour and is then kept constant at about 4500 N to expedite the bearing degradation till the experiment ended, as shown in Figure 5.14(b). Additionally, the mean static pre-load is visually indicated by three colors to ease further analysis. The measured peak dynamic load deviates from the set dynamic peak load and represents the dynamic nature of the random load, as shown in Figure 5.14(c). The measured condition monitoring signals, namely the vibration signal measured with the horizontal accelerometer *C*, the temperature signal measured at the position *T2* and the *U*-phase motor current signal are exemplarily depicted in Figure 5.15. As observed from Figure 5.15(a), the vibration RMS is almost constant from the onset of the experiment till about 2.7 h, where it then gradually increases. Approximately about 7.3 h, the vibration RMS exponentially increases till the end of experiment, but with a slight decrease at about 7.6 h. According to Williams et al. [278] the decrease in the vibration RMS is attributed to smoothing of the rough fault surface area as the rolling element roll over it, while the subsequent increase in the vibration RMS is attributed to further broadening of the fault area. As also seen in Figure 5.15(a), the RMS value of the vibration signal is not directly influenced by the varying static pre-load.

As shown in Figure 5.15(b), in the bearing wear-in period, the absolute temperature increases from the ambient temperature of approximately 18 °C to the bearing operating

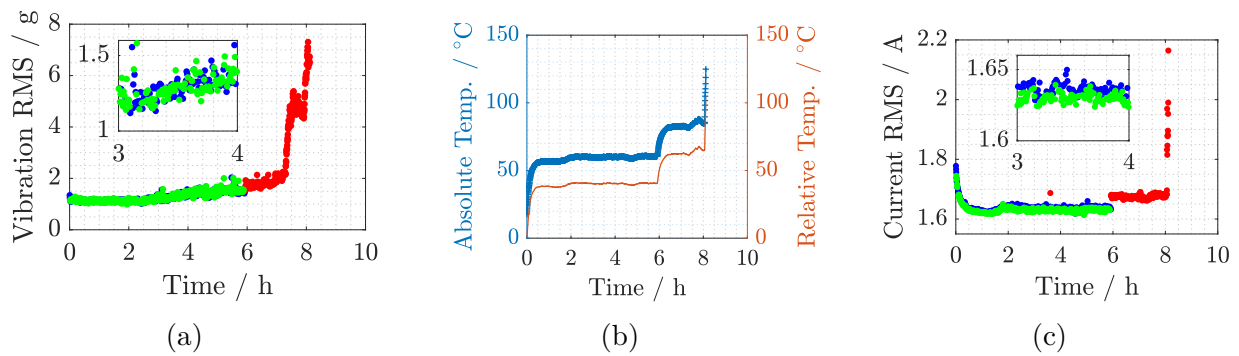


Figure 5.15: Condition monitoring signals of the run-to-failure experiment BS02: (a) vibration, (b) temperature and (c) motor current. The three colors in the vibration and motor current figures correlate to the three described load levels.

temperature of approximately 57°C, from the onset of the experiment till about 2.7 h. Thereafter, the temperature fluctuates significantly around the bearing operating temperature till approximately 6 h, where the temperature noticeably increases to 82°C, due to the increase in the static pre-load. The temperature further stabilizes until approximately about 7.3 h. Following that, the temperature slightly increases and then significantly increases up to 125°C at the end of the experiment, where the temperature-based threshold value T_{thresh} was exceeded. Additionally, as presented in Figure 5.15(b), the trend of the relative temperature suggests that the described effects on the absolute temperature is not as a result of the ambient temperature, but due to the variation in load amplitude and also attributed to degradation. Figure 5.15(c) shows that the RMS of the motor current decreases from approximately 1.8 A to 1.6 A in the bearing wear-in period, conversely to the temperature. Thereafter, the motor current remains almost constant at 1.6 A till approximately 6 h, subsequently the RMS of the motor current increases to 1.7 A due to the significant increase in static pre-load. Afterward, the motor current then fluctuates around the attained value of approximately 1.7 A till the third region is reached. Finally, the motor current progressively increases till approximately 2.2 A, when the experiment ended.

It can be summarized from the results of both preliminary experiments, that the vibration signal is directly influenced by the variation in the rotating speed, but not directly by the variation in static pre-load. Furthermore, the effects of the varying speed and moderate static pre-load are not significant in the temperature signal. However, an abrupt change in static pre-load of approximately 1500 N influences the temperature signal. The motor current signal is directly influenced by the variation in the rotating speed and static pre-load. Additionally, because the motor current signal does not show any consistent trend, as presented in the preceding preliminary analysis results, only the vibration and temperature signal are further considered in the following sections.

5.2.2 Influence of the operating conditions on the failure mode

The influence of the operating conditions on the failure mode is investigated via two experiments, namely experiment B02 and B03, subjected to similar distribution of varying operating conditions. Generally, in both experiments, the set rotating speed is randomly sampled from a stationary uniform distribution approximately within the interval [2400, 3000] rpm between measurement intervals. The set static pre-load and the set amplitude of the sinusoidal dynamic load are also randomly sampled from stationary uniform distributions within intervals [3300, 3800] N and [500, 1400] N, respectively. Additionally, in the experiment B02, a so-called warm-up phase of constant operating conditions is conducted for about 15 min before applying varying operating conditions. Specifically, the set rotating speed is 750 rpm, the set static pre-load is 2500 N and no dynamic load is exerted in the warm-up phase. A comparison of the distributions of the different operating conditions using violin plots is shown in Figure 5.16. As presented in the figure, both experiments have similar characteristics, such that the shapes and central tendency are similar. Particularly, the speed distribution resembles a uniform distribution and lies in the set values as previously described, as shown in Figure 5.16(a). Furthermore, the mean and median of experiment B03 are almost equal at approximately 2700 rpm, while the mean and median of experiment B02 are approximately 2600 rpm and 2700 rpm, respectively.

Figure 5.16(b) shows that the static load distribution resembles a normal distribution, al-

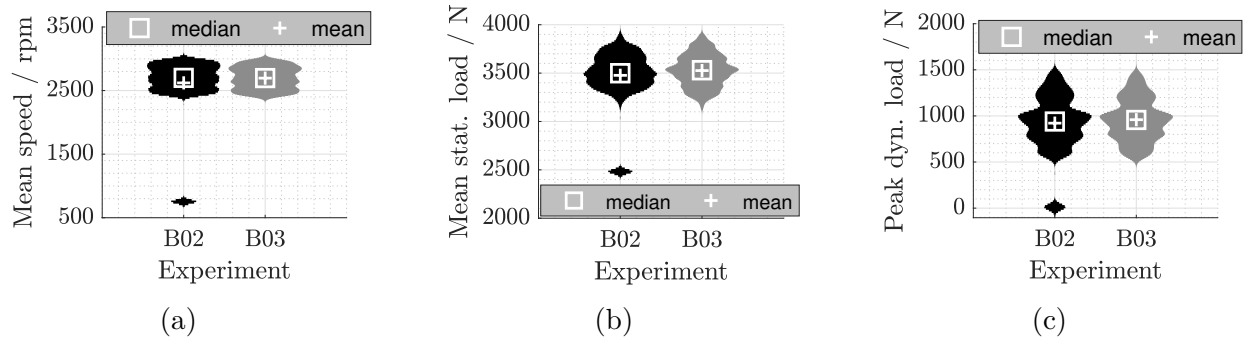


Figure 5.16: Comparison of operating condition distributions of: (a) varying speed, (b) varying static pre-load and (c) varying amplitude of a sinusoidal dynamic load for experiments B02 and B03

though it was sampled from a uniform distribution. This is attributed to the typical noisy nature of raw data and also to the central limit theorem. However, the values of the static pre-load lies within the range of set values, as can be inferred from Figure 5.16(b). Furthermore, the distributions of experiments B02 and B03 are almost similar with roughly symmetrical distributions, where the mode, mean and median are approximately 3500 N, as presented in Figure 5.16(b). As shown in Figure 5.16(c), although also sampled from a uniform distribution, the dynamic load distribution resembles a multi modal distribution and further reflects the dynamic nature of the raw signal. The similar distributions of both experiments are seen in Figure 5.16(c). Furthermore, they have similar local modes, mean and median of approximately 1000 N.

The acquired vibration and temperature signals relating to the conditions of the test bearings in both experiments are depicted in Figures 5.17(a) and 5.17(b), respectively. As shown in Figure 5.17(a) the vibration RMS of bearing B02 is about 0.2 g in the warm-up phase and increases abruptly to a mean of about 0.5 g at approximately 15 min into the experiment, as a result of a significant increase in the operating condition values. The vibration RMS varies around the mean value of approximately 0.5 g in the relatively long constant phase till about the fourth hour, where the transition in the faulty phase occurs. In the faulty phase, the intensity of the vibration reaches a peak of about 4.0 g and

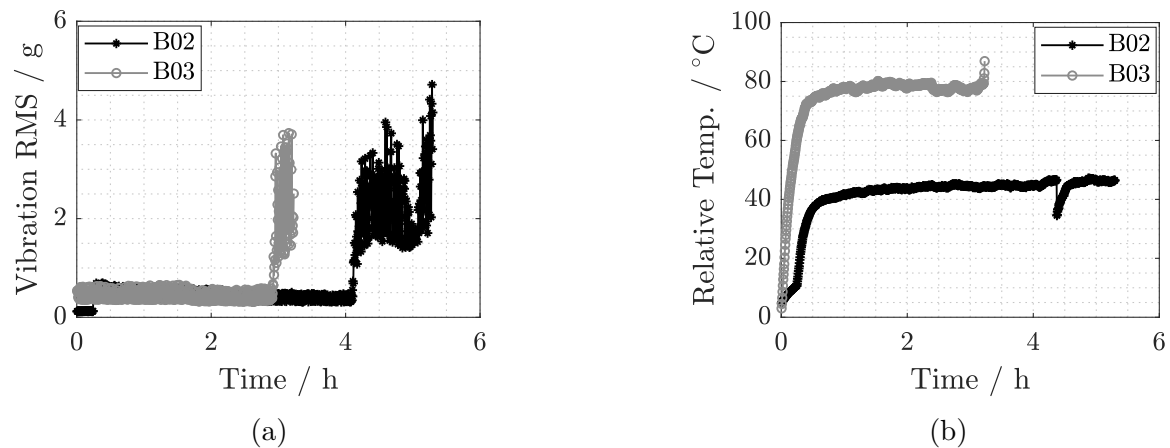


Figure 5.17: Comparison for experiments B02 and B03, considering the: (a) vibration RMS and (b) relative temperature.

decreases to about 1.7 g during almost one hour, due to initial spalling and subsequent smoothening of the sharp spall edges on the raceways during rolling of the balls on the raceway surfaces. Thereafter, as a result of continued bearing degradation, the vibration RMS further increases up to roughly 5.3 g, where the vibration-based threshold value is exceeded and the experiment ends.

Contrarily, for the experiment B03, there is no warm-up phase. Hence, the vibration RMS begins at approximately 0.5 g and fluctuates around this mean value up to about 2.9 h, where the vibration RMS then abruptly increases to about 4 g. The vibration RMS does not significantly increase further, but fluctuates around 2.7 g till the experiment ends. Figure 5.17(b) illustrates that the relative temperature of experiment B02 begins at approximately 5 °C and rises to approximately 11 °C. At about 15 min into the experiment, where the operating conditions are significantly increased, the slope of the bearing temperature also significantly changes and rises thereafter progressively up to about 40 °C from the temperature-rise phase to the steady-state operating temperature of the bearing. The relative temperature subsequently rises slightly about 7 °C more, after approximately four hours. It then slightly drops from 47 °C to about 35 °C at about 4.3 h, because the experiment was interrupted by the operator. After the experiment was resumed, the bearing temperature increases to the steady-state operating temperature of about 46 °C. Even at the end of experiment, due to increased bearing vibration, the bearing temperature does not significantly increase. Contrarily, for experiment B03, the three phases of the temperature curve can be inferred from Figure 5.17(b). In the first phase, the relative temperature increases from 3 °C to about 60 °C after approximately 15 min. In the second phase, the relative temperature fluctuates around the bearing operating temperature of approximately 78 °C. In the third phase from about 3.2 h, the relative temperature progressively increases till about 87 °C, where the temperature threshold limit is exceeded.

Figure 5.18 shows that the failure mode is spalling in the inner raceway of the test bearings after dismantling, in both experiments B02 and B03. Despite the width of the spall is similar for both bearings, the length of the spall is almost twice as large for the experiment B02 as for B03. However, as presented in Figure 5.17(a), the vibration RMS of the experiment B02 does not clearly reflect this severity, because its value is not twice as large as for experiment B03. Furthermore, Figure 5.17(b) shows that the steady-state relative temperature of experiment B03 is almost twice as large as the relative temperature of



Figure 5.18: Spalling on the inner raceways of the test bearing after the experiment: (a) B02 and (b) B03

experiment B02, which is contrary to the difference in fault extent. Although the similar operating conditions in both experiments resulted in inner race faults of different extents, it cannot be generalized that such operating conditions as described here always lead to solely inner race faults.

Although inner race faults are more frequent in rolling element bearings [288], outer race faults or a combination of inner race, outer race and rolling element faults can also occur. Specifically, under ideally pure radial loading on the bearing with a stationary outer ring and rotating inner ring, it is expected that faults occur concurrently in both the inner and outer race [223].

5.2.3 Influence of the operating conditions on the degradation curve

To actively counteract abrupt fault development, which predominantly occurred in the previous experiments, the operating conditions were substantially decreased during the onset of the faulty phase in experiments BD01 and BD02. The aim of the experiments is to shift the abrupt degradation to an incipient and gradual degradation, which is reflected in the condition monitoring signals and hence achieve at least a balanced split of healthy and unhealthy phase. In both experiments, the set rotating speed n_k is uniformly sampled from six distinct values, such that $n_k \in \{200 \cdot i : i \in \mathbb{N}, 8 < i < 13\}$ rpm. The set static pre-load is constant at 3.4 kN and 3 kN in experiments BD01 and BD02, respectively. Furthermore, the dynamic load is sinusoidal with a constant set amplitude of 1.1 kN and 1.5 kN in experiments BD01 and BD02, respectively. During the onset of the faulty phase, which is inferred from the measured vibration signal, the set rotating speed is decreased, such that $n_k \in \{100 \cdot i : i \in \mathbb{N}, 15 < i < 20\}$ rpm in both experiments. Furthermore, the set static pre-load is decreased by approximately 20% in both experiments, while the set dynamic load amplitude is decreased by approximately 33% and 45% in experiments BD01 and BD02, respectively.

The vibration RMS of the horizontal accelerometer C and relative temperature at the position $T2$ are depicted in Figure 5.19 for both experiments. As presented in Figure 5.19(a), the vibration RMS begins at an average of approximately 0.4 g and increases

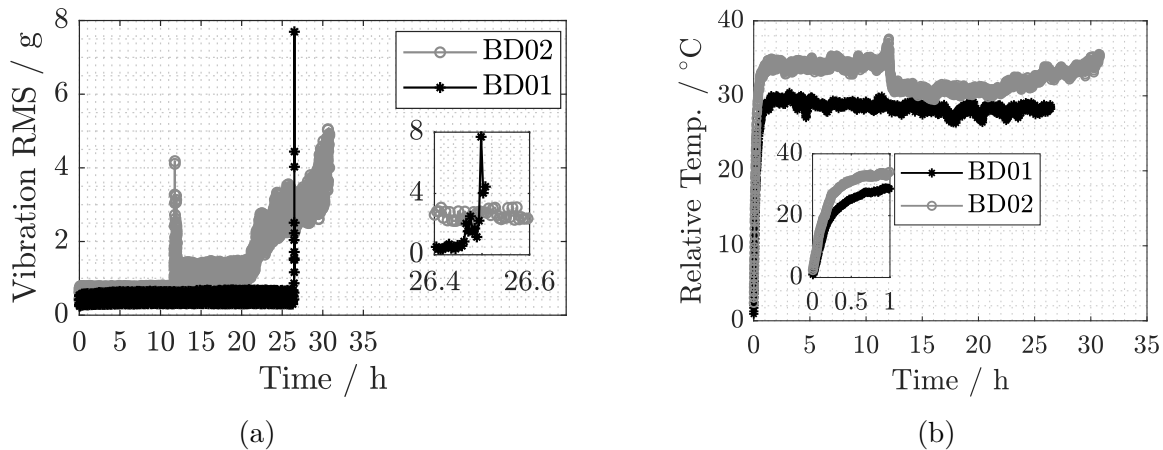


Figure 5.19: Comparison for experiments BD01 and BD02, considering the: (a) vibration RMS and (b) relative temperature

slightly to an average of approximately 0.6 g for experiment BD01. Furthermore, the vibration RMS fluctuates around the mean of approximately 0.6 g almost over entire the experiment till about 26.5 h, where it increases to a vibration RMS level of approximately 2.5 g, as depicted in the enlarged portion of the figure. Subsequently, the vibration RMS slightly decreases, as a result of the decrease in the operating conditions so as to promote incipient degradation. After about four subsequent measurements, the vibration RMS further increases up to 7.7 g due to progressive degradation, as shown in Figure 5.19(a). The experiment ended shortly thereafter, because the vibration-based threshold value was exceeded. Contrarily, Figure 5.19(b) illustrates that the relative temperature does not show any significant increase in the faulty phase for the experiment. After the experiment, the dismantled test bearing revealed a ball fault on a single rolling element.

Figure 5.19(a) shows that the vibration RMS fluctuates significantly around a mean of approximately 0.6 g till about 11.7 h for experiment BD02. Furthermore, the vibration RMS subsequently increases to a maximum of approximately 4.2 g followed by a short period of decrease and increase in the vibration RMS, due to smoothening and further degradation. As shown in Figure 5.19(a), at about the twelfth hour into the experiment, the vibration RMS decreases to approximately 1 g, due to the degradation-related significant decrease in the operating condition. For further eight hours, the vibration RMS stabilizes at approximately 1 g and gradually increases at about the 22 h. The vibration RMS further gradually increases till approximately 5 g, when the experiment ended at about 30.7 h. As presented in Figure 5.19(b), although the temperature does not significantly increase at the end of the experiment, it gradually increases by 5 °C at about the 22 h onward. Furthermore, it also reflects the decrease in the operating condition value, via a decrease in the temperature value by approximately 7 °C at about the twelfth hour. In this experiment, the fault mode of the test bearing is a multi-fault mode, that is, a combined ball, inner race and outer race spalling.

In summary, in the case of a ball fault, a significant reduction of the operating condition did not significantly influence the degradation curve, due to the strong impulse emanating from the ball contact on the inner ring, outer ring and cage [289]. Contrarily, inner and outer race fault tend to develop gradationally. Hence, a modification of the operating condition towards influencing the degradation curve tends to be effective. In any case, the resulting fault mode is not predetermined a priori, but initiate during the run-to-failure experiment. Thus, leading to additional challenge and accompanied uncertainty.

5.2.4 Overview of experiments

In the following, the experimental dataset acquired and published [290] in the course of this work is further investigated upon. An overview of the operating condition set values for the 17 experiments are presented in Table 5.1, while further information pertaining to the dataset, such as the threshold criteria and number of measurements per dataset can be found in Table 5, in the Appendix. In some experiments, the three operating conditions, namely, the rotating speed, the static pre-load and the dynamic load amplitude are varied within the specified intervals. In other experiments, two of the three operating conditions are kept constant and one is varied. Furthermore, the failure modes are diverse, such that single faults, such as inner race faults and multi-mode fault in several bearing components, occurred in the run-to-failure experiments.

Table 5.1: Set range of operating conditions for experiments

Experiment	Rotating Speed / rpm	Static Load Amplitude / N	Dynamic Load Amplitude / N	Dynamic Load Excitation	Defect
B01	[2400, 3000]	[3300, 3800]	[500, 1400]	Sinusoidal	IR, OR, B
B02 - B03	[2400, 3000]	[3300, 3800]	[500, 1400]	Sinusoidal	IR
B04	[1500, 3000]	[2500, 3800]	[500, 1500]	Sinusoidal	IR, B
B05	[1500, 3000]	[2500, 3800]	[500, 1500]	Sinusoidal	B
B06	[1500, 3600]	[3300, 3800]	[750, 1700]	Sinusoidal	IR, B
B07	[1500, 3000]	[3250, 4000]	[250, 750]	Gaussian noise	IR
B08	[1500, 3000]	[3250, 4000]	[500, 1000]	Gaussian noise	IR, B
B09	[1500, 3000]	[2500, 3800]	[750, 1000]	Gaussian noise	OR, B
B10	[1500, 2700]	[2000, 3250]	[750, 1500]	Gaussian noise	OR, B
B11	[1500, 2700]	3000	1000	Gaussian noise	IR, OR, B
B12 - B13	[1500, 2700]	3000	1000	Gaussian noise	IR, B
B14	2700	2500	[700, 1500]	Gaussian noise	B
B15	2700	2500	[700, 1500]	Gaussian noise	IR, B
B16	2700	2500	[1000, 1500]	Gaussian noise	IR, OR, B
B17	2700	[1500, 2500]	1500	Gaussian noise	B

Where IR denotes inner race defect(s), OR denotes outer race defect(s) and B denotes rolling element defect(s).

5.3 Fault detection approaches

In the following, the described methodology is elaborated based on the bearing run-to-failure datasets.

5.3.1 Approach A: Signal analysis techniques

In this section, the time-frequency domain analysis method toward fault detection is investigated. Furthermore, since the rotating speed is synchronously measured with the vibration signal, the order analysis of the horizontal accelerometer C is considered. The experiment B02, previously analyzed in Section 5.2.2 is exemplarily analyzed here. The resulting spectrogram of the order analysis is shown in Figure 5.20(a). It can be partitioned vertically in three zones characterized by the amplitude intensity. The first zone, with relatively low amplitudes, correlates with the warm-up phase and ends approximately at about 15 min. The second zone, with most of the bearing life time, is mostly characterized by whole multiples of the shaft rotating frequency and also by the ball pass frequency outer (BPFO). Although the BPFO is said to correspond to a faulty outer race, this frequency component also typically occurs for a healthy bearing due to the loading situation and bearing clearance [291], [292], [293]. Furthermore, the bearing can be considered to be in an healthy state in the second zone, because till approximately the fourth hour side-bands are not present. Moreover, in the third zone, where the bearing significantly degrades, whole multiples of the ball pass frequency inner (BPFI) and side bands are visible. Furthermore, the amplitude intensity also significantly increases.

After the experiment, the dismantled bearing revealed an inner race fault. The maximum amplitude considering up to the 25th order is depicted in Figure 5.20(b), over the entire measurement time. The three zones as previously described are also seen in Figure 5.20(b), that is, the first zone with almost 0 g, the second zone that fluctuates approximately about [0.1, 0.6] g and the third zone with a maximum peak amplitude of almost 6.6 g. The entire trend is similar to the vibration RMS, but different in the third zone, because the maximum peak occurs here almost about an hour before bearing failure, as opposed to the

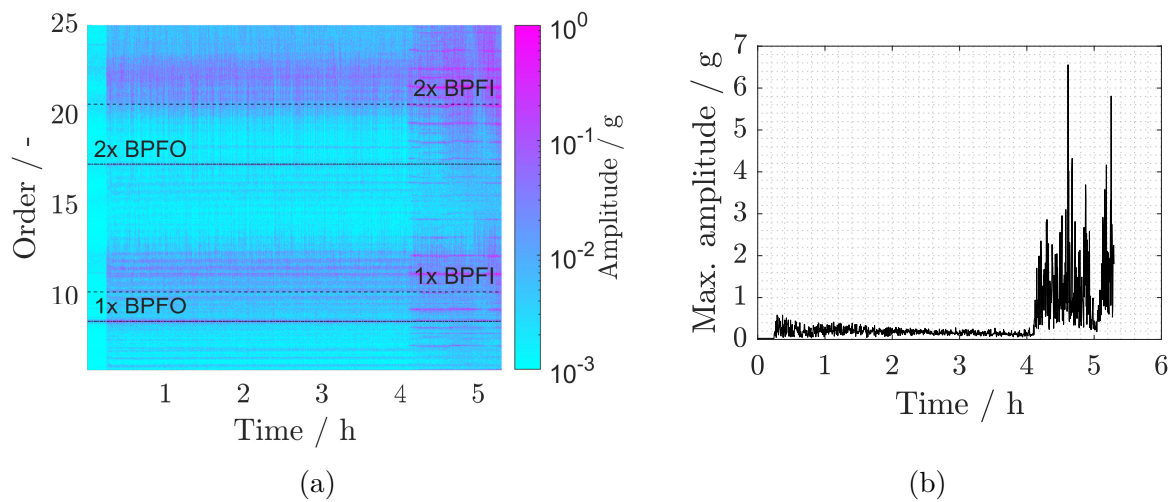


Figure 5.20: Time-frequency analysis of B02 via (a) order analysis up to 25th order and (b) corresponding maximum amplitude

vibration RMS, where the maximum peak occurs shortly before the end of the experiment. Although the order analysis normalizes the frequencies to the rotating frequency, the effects of loading situation still dominantly influences the analysis, as exemplarily seen in the warm-up phase.

A further analysis is the spectrogram analysis in the high frequency region above 10 kHz. For some experiments with harmonic dynamic load, amplitude peaks were found to depict characteristic trends at about 15 kHz, 35 kHz and 55 kHz, such that the frequency is decreased over time. Furthermore, the trend is almost mirrored at about 20 kHz, 40 kHz and 60 kHz. An excerpt of the spectrogram in the frequency range [40, 64] kHz is depicted in Figure 5.21(a). Although the characteristic peaks might be related to the resonance frequencies of the bearing components or the bearing clearance, it could not be validated, because the experimental determination of the bearing components' resonance frequencies and clearance was not feasible, due to the dimension and composition of the test bearing. The corresponding frequency at maximum amplitude in the range [50, 55] kHz is exemplarily depicted in Figure 5.21(b). The frequency decreases rapidly from about 54 kHz with approximately 700 Hz in the first hour, as shown in Figure 5.21(b). Although the frequency further decreases, the slope becomes shallow at about 1.6 h and shallower at about 4.3 h onward. Due to the dynamic nature of the random dynamic load and the relative low amplitude in the high frequency region, the peaks are buried in noise, in the experiments with random dynamic load. Thus, similar decreasing trend in the frequency domain could not be extracted. A further drawback is that, to detect a fault, a domain expert is required for analysis or a threshold needs to be set, under consideration of sufficient data of similar trends.

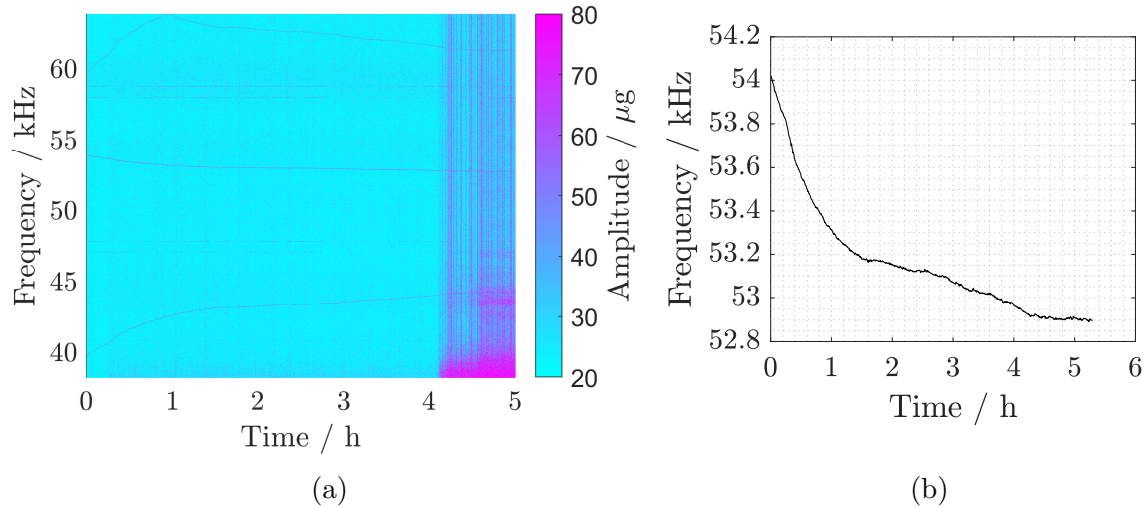


Figure 5.21: Time-frequency analysis of B02 via (a) spectrogram in the frequency range [40, 64] kHz and (b) corresponding frequency at maximum amplitude in the range [50, 55] kHz

5.3.2 Approach B: Condition-invariant feature engineering

The previously described feature engineering techniques are exemplarily employed here on the vibration data acquired with the horizontal accelerometers A and C. Specifically, several classical time-domain features, features from the HCTSA package and features from the EMG toolbox are extracted, as described in [137]. Furthermore, the filter-based feature selection technique, namely monotonicity, is subsequently employed to identify a subset of features, that at best indicate the continuous degradation of the bearings, irrespective of the time-varying operating conditions. However, the application of the feature engineering techniques per experiment yielded distinct top features for the different experiments, which limits the transferability of a set of features over all experiments. Hence, to find a set of best features over all experiments, a subset of 100 features per horizontal accelerometers A and C, including top features from the individual experiments are further investigated. The first top-ranked feature identified is the so-called `MD_pNN_raw.pnn70` feature extracted from the horizontal accelerometer A with the HCTSA package [216]. The `MD_pNN_raw.pnn70` feature derived from the proportion of normal-to-normal heart beats (pNN) is one of the mostly employed time-domain statistic from the biomedical domain to measure heart rate variability [294]. It is the mean number of consecutive difference that a signal exceeds a predefined threshold value [294]. Figure 5.22 exemplarily depicts the `MD_pNN_raw.pnn70` feature extracted for three experiments, considering different influences of the operating conditions.

Experiment BS01 is the preliminary experiment described in section 5.2.1, which considers the effect of varying speed, while the static pre-load and sinusoidal dynamic load amplitudes are constant. Experiment B02 considers varying speed, varying amplitudes of the static pre-load and sinusoidal dynamic load amplitude, while experiment B17 considers constant speed, varying static pre-load and random dynamic load amplitude. Figure 5.22 demonstrates that the feature allows at least a clear distinction between healthy and unhealthy instances. For all three experiments, the feature value is almost zero in the healthy state and although it distinctively increases in the faulty state, the trends are not comparable. Notwithstanding, two possibilities have been identified to partition the instances in the healthy and unhealthy instances. The partition point, which is typically

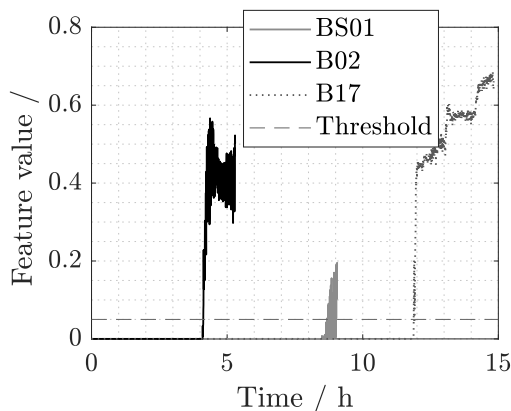


Figure 5.22: `MD_pNN_raw.pnn70` feature exemplarily extracted for three experiments, while considering an expert-defined threshold

Table 5.2: Derived transition times based on the `MD_pNN_raw.pnn70` feature

Experiment	Transition time [h]	
	Threshold	Cluster
BS01	8.68	8.73
B02	4.12	4.13
B17	11.88	11.95

referred to as transition time [169], can be derived via an expert-defined threshold value or via a binary cluster algorithm. Based on sensitivity analysis, the threshold value is set to 0.05. As shown in Figure 5.22, the threshold value is conservative and for experiment BS01 that took approximately nine hours, maybe too conservative. However, this is due to the signal that does not depict a continuous degradation trend in the faulty state, but depict a series of fluctuations, which is also common to other experiments.

K-means clustering algorithm with two clusters is also employed to partition each experiment in healthy and faulty instances. The transition point is then defined as the partition time instance, where the cluster value deviates from its initial value [169]. As presented in Table 5.2, although the clustering method tends to provide a later transition time than the expert-defined threshold method, it is an unsupervised method that does not require any domain knowledge. Furthermore, the difference between both estimated transition times might be insignificant in some applications.

5.3.3 Approach C: Condition-centered data or feature scaling

Scaling condition monitoring data or derived features thereof requires the operating condition information. In this section, the RMS feature extracted from the horizontal accelerometer C is exemplarily considered for two experiments subjected to different profiles of varying operating conditions. One of the experiments, namely the preliminary experiment BS01, considers varying speed, constant static pre-load and constant sinusoidal dynamic load amplitude. Thus, feature scaling is performed with respect to the speed. The other experiment, namely experiment B02, considers varying speed, varying static pre-load and varying sinusoidal dynamic load amplitudes. Therefore, the RMS feature extracted from the horizontal accelerometer C is scaled independently to the three operating condition variables.

Figure 5.23 presents the results of the feature scaling for both experiments in comparison to the RMS values without scaling. In the figure, the RMS values are scaled with respect to the mean speed $\bar{n}$, with respect to the mean static pre-load $\bar{F}_{stat}$ and with respect to the mean sinusoidal dynamic load amplitude $\bar{F}_{amp}$. Due to the higher range of values of the operating conditions as compared to RMS feature values, the resulting scaled values are typically very small. As shown in the top sub-figure of Figure 5.23(a), the effects of the speed are significant in the RMS feature values, as represented by the different horizontal regions in the zoomed portion. In the bottom sub-figure of Figure 5.23(a), where the RMS feature values are scaled to the mean speed, the entire curve seems less noisy. Furthermore, the effects of the speed is minimized, even if not completely removed, as better seen in the enlarged portion of the figure. As presented in the top sub-figure of Figure 5.23(b), the RMS feature values clearly reflect the effects of the warm-up phase till approximately 15 min into the experiment, where the operating conditions are significantly increased. As shown in the bottom sub-figure of Figure 5.23(b), the speed-scaling of the RMS feature values considerably minimizes the effects of the warm-up phase. However, the degradation part is not noticeably influenced by the speed-scaling. As opposed to the speed-scaling, the scaling with respect to the static pre-load has a negligible effect on the RMS feature values, while the scaling with respect to the sinusoidal dynamic load amplitude has an inverse effect in the warm-up phase and a rather negative effect in the faulty state. The inverse effect in the warm-up phase is attributed to the division by a relative small values as compared to the large values of the dynamic load after the warm-up phase. Furthermore, the negative effect in the faulty state is attributed to local peaks

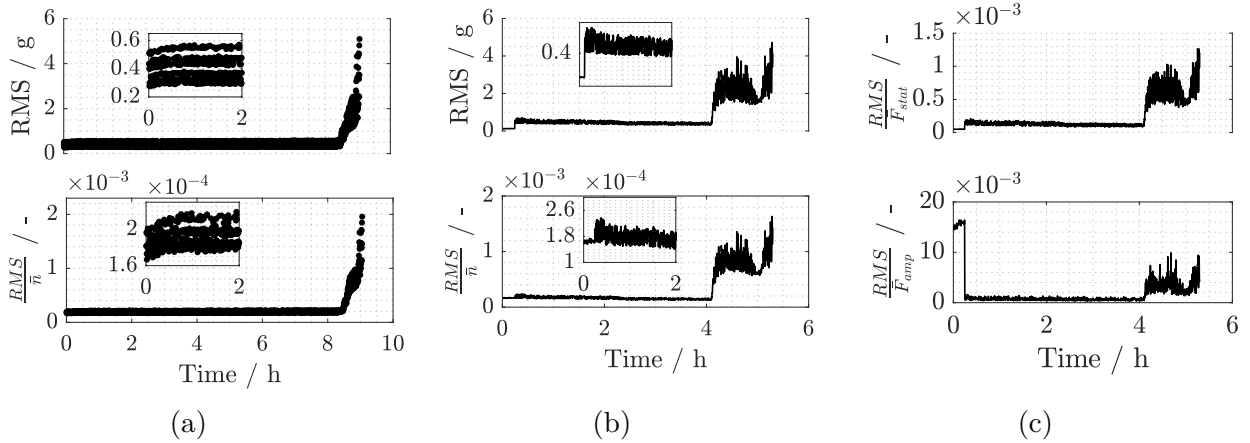


Figure 5.23: Feature scaling based on: (a) varying speed of experiment BS01, (b) varying speed of experiment B02, and corresponding (c) varying static pre-load and varying sinusoidal dynamic load amplitude

in the RMS feature intensified by the dynamic load amplitude.

The results obtained from scaling the features with respect to the operating conditions correspond with the findings obtained in Section 5.2.1, where the vibration signal is directly influenced by the variation in the rotating speed, but to some extent not directly influenced by the variation in the load amplitudes. Thus, as described in Section 5.3.2, an expert-based threshold or a cluster algorithm can then be employed to detect a fault, based on the speed-scaled feature values.

5.3.4 Approach D: Condition-specific feature standardization and feature segmentation

Based on the assumption that the operating condition or resulting modes of operation are discrete with a moderate number of clusters, then the condition monitoring data can be standardized or segmented according to each discrete operating condition profile. The three operating conditions of the ball bearing application example are randomly sampled from a stationary uniform distribution within specified operating ranges between measurement intervals. Hence, the operating conditions are not discrete, but bounded. However, to investigate if the resulting modes of operation are discrete with moderate number of clusters, a pairwise scatter diagram of the three operating conditions and the RMS of the vibration data measured with the horizontal accelerometer C are regarded exemplarily. Specifically during the analysis of experiment B02, the warm-up phase is excluded, for better visualization. The resulting pairwise scatter diagram is depicted in Figure 5.24, where the three groups designated by different markers are arbitrarily clustered around the static pre-load to facilitate data exploration.

As depicted in Figure 5.24, the rotating speed and the sinusoidal dynamic load amplitude do not form distinguishable clusters. In contrast, because the data has been clustered around the static pre-load, three distinguishable clusters of low, medium and high load regions can be seen in Figure 5.24. As also can be inferred from the third row of Figure 5.24, the combination of both the sinusoidal dynamic load amplitude and static pre-load have a lower bound at approximately 4.3 kN and an upper bound at approximately 4.9 kN. Furthermore, the sinusoidal dynamic load amplitude and static pre-load possess

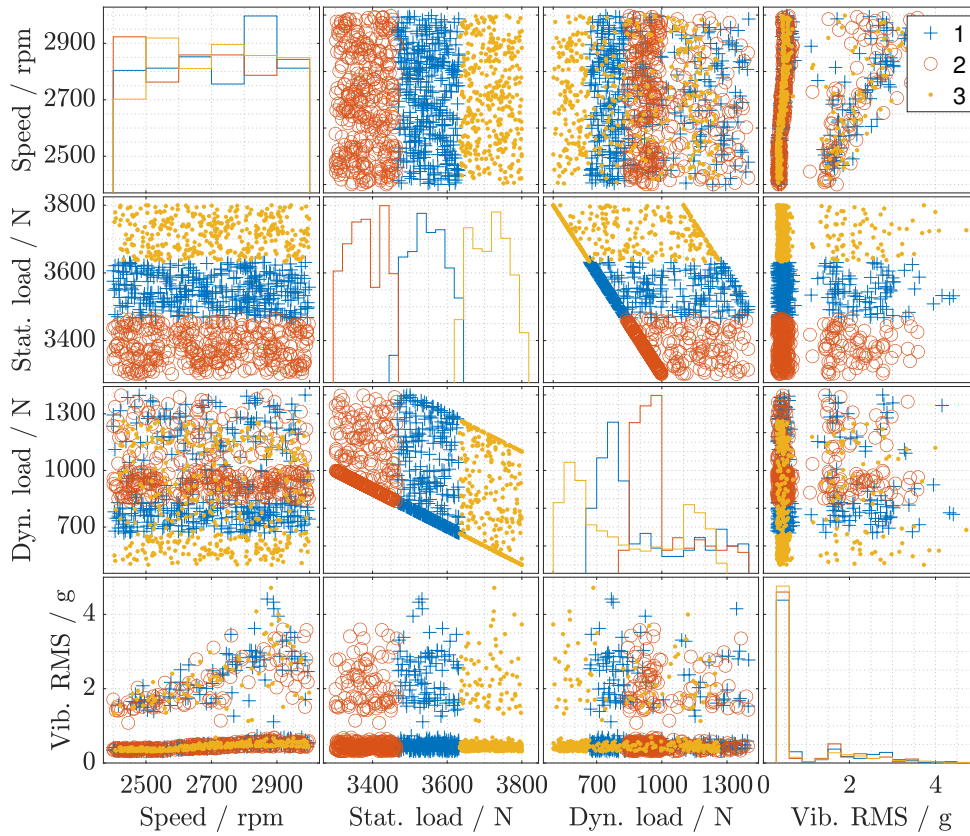


Figure 5.24: Pairwise scatter diagram involving the operating conditions and the vibration RMS

a linear relationship, such that an increase in the sinusoidal dynamic load amplitude is generally accompanied by a decrease in the static pre-load, and inversely, to maintain the fixed bounds at $[4.3, 4.9]$ kN. As can be inferred from the last row of Figure 5.24, for this experiment, a direct relationship does not appear to exist between the vibration RMS and the load amplitudes. Contrarily, as can be inferred from the first column of the last row, the vibration RMS slightly tends to increase with the speed in the healthy state, that is, for vibration RMS values approximately less than 0.8 g and significantly increases with the speed in the faulty state, where the vibration RMS values is approximately greater than 1 g. However, because the resulting clusters are also not distinctively separable, feature standardization per mode of operation does not counter for influence of the operating condition and feature segmentation is not feasible.

5.3.5 Approach E: Embedding in the modeling phase

In fault detection, the measured operating condition information can be incorporated during model development to consider the influence of time-varying operating condition. Binary clustering of the vibration data measured with the horizontal accelerometer C for the experiment B02 is exemplarily considered here to investigate the influence of embedding the measured operating condition information. In particular, four levels of integrating the three operation condition information are considered. In all cases, the

k-means clustering algorithm with two clusters is considered. In the first level, the vibration RMS is employed as input to the k-means model, but not the operating condition information. The resulting cluster is shown in the top sub-figure of Figure 5.25(a). In this case, the transition time is approximately 4.13 h, irrespective of the warm-up phase and the time-varying operating condition.

Alternatively, in addition to the vibration RMS, the mean speed, the static pre-load and sinusoidal load amplitude are employed as input, in the second case. As shown in the bottom sub-figure of Figure 5.25(a), the transition time is defined by the warm-up phase and approximately about 15 min. Contrary to the expectation, incorporating the three operating condition information did not compensate their influence, although the resulting cluster was not entirely stable. With the knowledge that the vibration RMS is significantly influenced by the mean speed, only the speed information is considered in the third case. As presented in the top sub-figure of Figure 5.25(b), incorporating only the speed information leads to a later transition time at approximately about five hours into the experiment, than incorporating all available operating condition information. However, the lead time before failure is shorter in comparison to the derived transition time considering only the vibration RMS. As shown in the bottom sub-figure of Figure 5.25(b), exemplarily incorporating only the sinusoidal dynamic load amplitude does not lead to two distinct partitions of healthy and unhealthy instances, because a direct relationship does not exist between the vibration RMS and the load amplitudes.

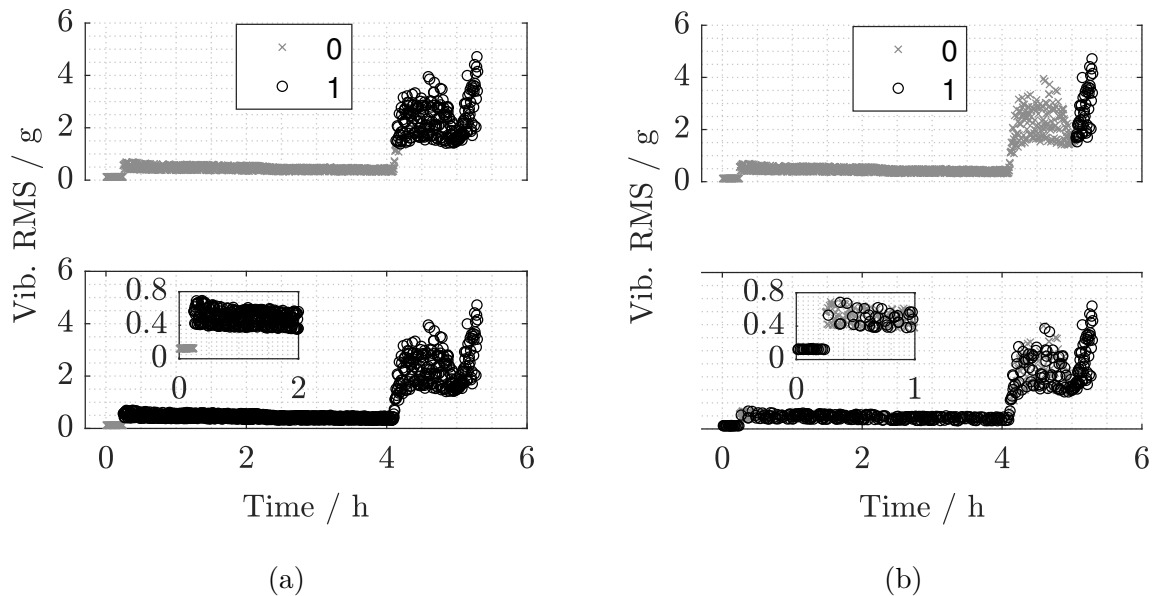


Figure 5.25: Binary clustering of the vibration RMS with different levels of additional input information (a) top: without the operating condition; bottom: with all operating condition (b) top: with the speed; bottom: with the dynamic load

5.4 Performance evaluation of the fault detection approaches

As discussed in Section 2.6, the evaluation of a diagnostic or prognostic method requires some knowledge about the ground truth. In the considered application example, the ground truth health state of the bearings is not directly observed over the lifetime, but inferred from the measured condition monitoring signals. Furthermore, the declaration of a hard boundary between healthy and faulty depends on the signal evaluation and expert intuition. Thus, the fault detection methods cannot be evaluated by the confusion matrix or metrics derived thereof. Nonetheless, the transition points and corresponding transition times are provided in Table 5.3 in comparison to a similar study by Javanmardi et al. [169], where different feature sets and number of features were considered to train a binary k-means algorithm.

In Table 5.3, approach J is the reference approach by Javanmardi et al. [169]. Approach B solely utilizes the MD_pNN_raw.pnn70 feature extracted from the vibration signal obtained with the horizontal accelerometer A, as presented in Section 5.3.2. In approach C, the vibration RMS of the horizontal accelerometer A is scaled with respect to the mean speed, while the mean speed is incorporated alongside the vibration RMS of the hori-

Table 5.3: Comparison of the transition times with the results obtained with approach J [169] as a reference approach

Bearing	Approach J		Approach B		Approach C		Approach E	
	Transition point		Transition point		Transition point		Transition point	
	Index / -	Time / h	Index / -	Time / h	Index / -	Time / h	Index / -	Time / h
B01	208	1.09	212	1.11	62	0.32	51	0.27
B02	998	4.74	870	4.13	870	4.13	51	0.24
B03	562	2.96	556	2.92	556	2.92	3	0.02
B04	452	2.25	452	2.25	444	2.21	50	0.25
B05	560	2.78	556	2.76	547	2.71	49	0.24
B06	12853	63.56	12838	63.48	9878	48.85	219	1.08
B07	6205	36.95	6204	36.94	52	0.31	50	0.30
B08	1219	7.13	1213	7.09	1213	7.09	50	0.29
B09	253	1.77	313	2.19	256	1.79	50	0.35
B10	2679	117.29	2679	117.29	2679	117.29	52	2.28
B11	931	4.61	719	3.56	978	4.84	3	0.01
B12	154	0.51	67	0.22	60	0.20	3	0.01
B13	18022	59.89	17705	58.84	17705	58.8	3	0.01
B14	12050	40.10	11956	39.79	11959	39.80	9209	30.65
B15	21051	70.06	21051	70.06	21611	71.92	647	2.15
B16	5400	17.95	5409	17.98	5414	18.00	5890	19.58
B17	2323	12.60	2202	11.95	2189	11.88	224	1.22

zontal accelerometer A to train a binary k-means algorithm, in approach E. Approaches A and D are not considered here due to their shortcomings. As can be inferred from Table 5.3, the transition points are similar for approach J and approach B in most of the bearings, except for bearings B11 and B13 with significant deviations of at least an hour. Except for bearings B14 and B16, approach E substantially deviates from the other approaches, such that it mostly underestimates the transition points. Thus, approach E predominantly underperforms the other approaches. However, approach C slightly underperforms approaches J and B, because it provides comparable transition points for most of the bearings, except for bearings B01, B06 and B07. To visualize the transition points, the vibration RMS of the horizontal accelerometer C of bearings B10 and B11 are exemplarily depicted in Figures 5.26(a) and 5.26(b), respectively. As shown in Figure 5.26(a) for the bearing B10, the transition time estimated by approach E is within the healthy state, and specifically shortly after the warm-up phase. Contrarily, the transition times estimated by the other approaches are equal and denote the onset of the faulty state, based on the underlying vibration signal. As shown in Figure 5.26(b) for the bearing B11, the transition time estimated by approach E is also within the healthy state, shortly after the experiment began. However, the transition times estimated by the other models are within the faulty state, because up to approximately 3.6 h and beyond, the operating condition distribution was not changed and specifically there was no warm-up phase for bearing B11. However, the transition time estimated by approach B can be considered the onset of the faulty state as opposed to approaches J and C that estimate a later transition time, based on the underlying vibration signal.

Considering the presented results for the transition time estimation and consequently the fault detection methods, it can be summarized that approach B, namely the binary clustering method based on at least a condition-invariant feature is capable of reliably discriminating between an healthy and unhealthy state, even under time-varying operating conditions. Contrarily, incorporating the speed information as an additional information in the modeling phase can result in an adverse effect, as in approach E. Scaling of the RMS feature value with respect to the speed, as in approach C provides comparable transition

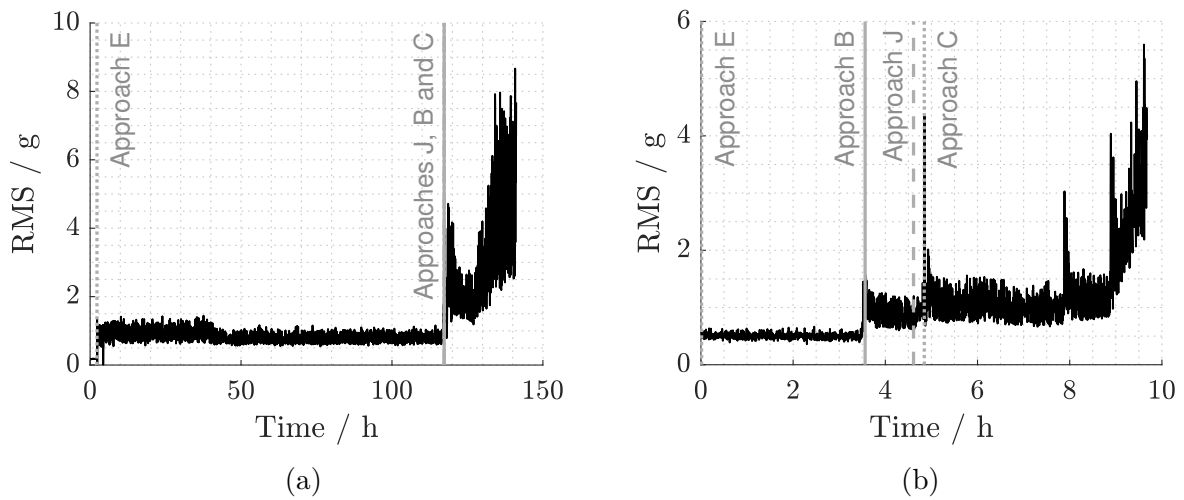


Figure 5.26: Comparison of the transition times estimated via approaches B, C and E as compared to approach J [169], while considering the vibration RMS of the horizontal accelerometer A of (a) bearing B10 and (b) bearing B11

points for most of the bearings, while the reference approach J does not primarily consider the operating condition, but requires many features.

5.5 Prognostics approaches

This section focuses on the possibilities of considering time-varying operating conditions during point estimation of the RUL of ball bearings. A first approach utilizes the empirical bearing basic rating life L_{10} , which directly incorporates the dynamic bearing load and the rotating speed, while a second approach utilizes machine learning technique to learn the underlying relationship between the condition monitoring data, the operating conditions and the RUL, based on the bearing run-to-failure datasets. For a fair comparison over all datasets, approach A and specifically the features derived from the time-frequency spectrum in the frequency range [50, 55] kHz is not further considered here, because as discussed in Section 5.3.1, the features were characteristically significant only for the first six experiments with sinusoidal dynamic load excitation. The other approaches of considering time-varying conditions, namely, condition-invariant features (approach B), feature scaling (approach C) and feature standardization (approach D) are separately investigated as possible input to a machine learning model. Furthermore, in approach E, the operating conditions are also investigated as an additional input to a machine learning model, similar to the work of Javanmardi et al. [169].

5.5.1 RUL estimation based on an empirical formula

The modified rating life, as defined in equation 5.1 can directly be employed here, because the operating condition information are available and although the operating conditions vary over the entire life time, they are almost constant per measurement. Furthermore, the basic dynamic radial load rating C of the test bearing is available from the bearing data sheet, namely, 4 kN. The dynamic equivalent load P per measurement is the sum of the average static pre-load and dynamic load amplitude and the exponent p is three for ball bearings. Based on the hypothesis that the RUL linearly decreases over the entire life time, the RUL in hours at instance k can be formulated as in Equation 5.8.

$$RUL_k = \left(a_{1,k} \cdot a_{ISO,k} \cdot L_{10,k} \cdot \frac{10^6}{60 \cdot n_k} \right) - t_k, \quad (5.8)$$

where $a_{1,k}$ and $a_{ISO,k}$ denote the modification factors for reliability and condition-related factors, respectively, $L_{10,k}$ denotes the basic rating life, n_k is the rotating speed and t_k denotes the elapsed time in hours at instance k . The dynamic equivalent load P and the rotating speed n are depicted in Figure 5.27(a) exemplarily for the experiment B02. As can be seen in Figure 5.27(a), in the warm-up phase, the equivalent dynamic load is approximately 2.5 kN and the rotating speed is approximately 750 rpm. After the warm-up phase, the equivalent dynamic load and the rotating speed are not constant, but vary within a specified interval. Under consideration of the equivalent dynamic load and the rotating speed, but while neglecting the modification factors, namely, for $a_1 = 1$ and for $a_{ISO} = 1$, the estimated RUL based on Equation 5.8 is depicted in Figure 5.27(b). As can be inferred from Figure 5.27(b), in the warm-up phase, the RUL is significantly overestimated. The overestimation of the RUL is mainly attributed to the fact that the

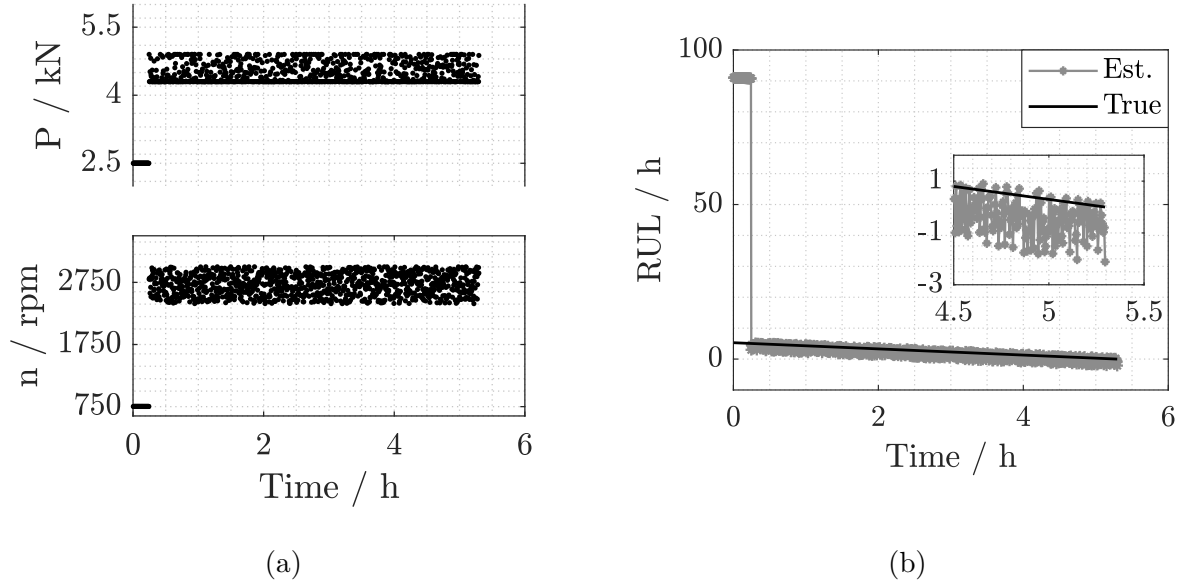


Figure 5.27: (a) Dynamic equivalent load P and rotating speed n for the experiment B02 and (b) corresponding RUL based on the rating life for $a_1 = a_{ISO} = 1$

theoretical basic rating life primarily considers the current operating conditions at instance k to estimate the current RUL. Furthermore, it is attributed to the relatively low load of about 40% less than the dynamic load rating of the test bearing and a rotating speed of approximately 95% less than the bearing limiting speed. After the warm-up phase, where the operating conditions are significantly increased to accelerate the degradation of the bearing, the estimated RUL reflects a significant decrease in the bearing life. However, the RUL is mostly underestimated and possess a large variance, due to the nature of the operating conditions. Moreover, the underestimation after the warm-up phase is attributed to the fact that as opposed to this experiment with a pronounced fault distributed over the surface of inner raceway, the theoretical basic rating life is valid up to fatigue initiation, where a spall becomes evident in at least one of the bearing components. According to the ISO 281 standard [226], the a_{ISO} modification factor considers the lubrication and contamination levels of bearings to improve on the accuracy of the estimated bearing life. Since the test bearings are grease-lubricated, lubricated-for-life and sealed, the raceway are deemed free from contaminants. However, since the lubrication level could not be precisely determined, a modification factor $\tilde{a}_{ISO}$ is derived from the bearing temperature. The relative temperature T_2 is depicted in Figure 5.28(a) exemplarily for six representative experiments over the entire lifetime.

Although the temperature curves possess similar trends, the bearings possess wide variability in their lifetimes. Furthermore, as can be seen in Figure 5.28(a), bearings B13 and B15 with an operating temperature under 40°C tend to possess a longer lifetime as opposed to bearings B01 and B03 with an operating temperature above 60°C , which have relatively low lifetimes. Thus, based on the temperature measurements, the modification factor $\tilde{a}_{ISO}$ is derived according to the pseudo code in Algorithm 1. Specifically, $\tilde{a}_{ISO} = 1$ in the first temperature region, where the bearing temperature rises from the ambient temperature to the operating temperature or an equilibrium temperature T_{eq} . During the experiments, the equilibrium temperature T_{eq} is reached after approximately an hour into the experiment.

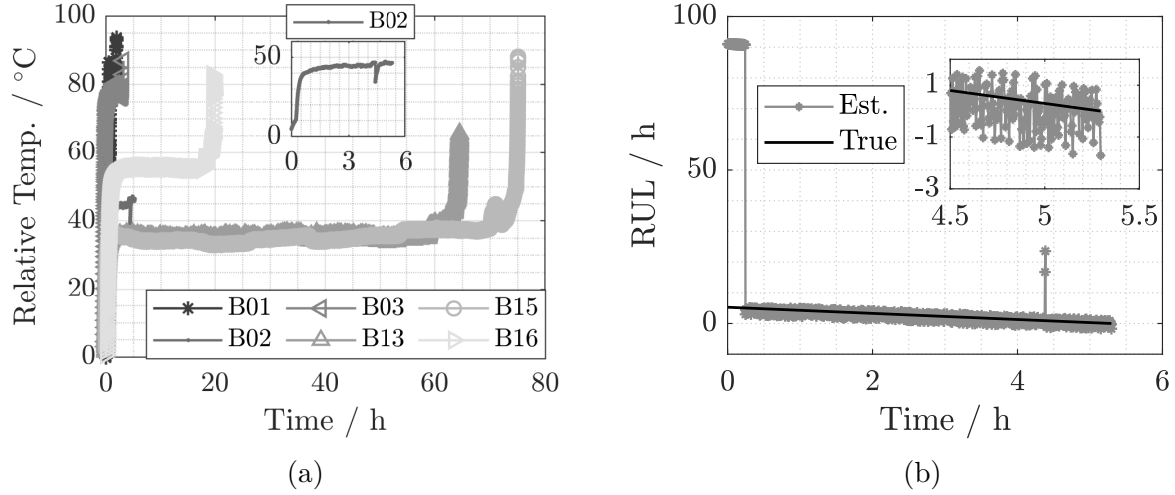


Figure 5.28: (a) Temperature measurements over the entire lifetime for some experiments
(b) RUL for experiment B02 based on the modified rating life considering the temperature measurements in the a_{ISO} modification factor

The modification factor $\tilde{a}_{ISO}$ is between five and seven, where the the temperature lies between 30°C and 37°C, to account for bearings with longer lifetimes. Furthermore, to account for bearings with an operating temperature above 60°C and comparative shorter lifetimes, the modification factor $\tilde{a}_{ISO}$ is between 0.3 and 0.5. For bearings that lie between both extremes and with an operating temperature between 37°C and 60°C, $\tilde{a}_{ISO}$ is between 1 and 1.3. Under consideration of the temperature-derived modification factor $\tilde{a}_{ISO}$, the estimated RUL is depicted in Figure 5.28(b), exemplarily for experiment B02. Particularly for the experiment B02, the lifetime is modified by the factor $1 \leq \tilde{a}_{ISO} \leq 1.3$, such that the estimated RUL is generally minimally improved, as shown in Figure 5.28(b). However, as also shown in Figure 5.28(b), the estimated RUL is significantly overestimated around about 4.5 h, due to the considerable decline in the bearing temperature. Although such decline in temperature could be as a result of improved lubrication of the raceways, it resulted from interruption of the experiment in this case.

Algorithm 1 Derivation of lifetime modification factor $\tilde{a}_{ISO}$ based on temperature measurements

```

if ( $T_k < T_{eq}$ )  $\vee$  ( $T_k \leq 30^\circ\text{C}$ ) then
     $\tilde{a}_{ISO,k} = 1$ ;
else if ( $30^\circ\text{C} < T_k < 37^\circ\text{C}$ ) then
     $\tilde{a}_{ISO,k} = [5, 7]$ ;
else if ( $37^\circ\text{C} < T_k < 60^\circ\text{C}$ ) then
     $\tilde{a}_{ISO,k} = [1, 1.3]$ ;
else if ( $T_k > 60^\circ\text{C}$ ) then
     $\tilde{a}_{ISO,k} = [0.3, 0.5]$ ;

```

Although the bearing rating life L_{10} provides a theoretical estimation of the RUL, it lacks the possibility of directly considering the monitored condition monitoring signals, such as the vibration signals. Furthermore, although the modified rating life via the consideration of the temperature measurement provides some improvements to the RUL estimation, it

does not converge towards the true EoL.

5.5.2 RUL estimation via machine learning regression

In the following, a supervised machine learning model for regression task, that is, for estimating the RUL as a continuous variable, is considered. As argued in the supporting publication [169], any machine learning model for regression task is applicable. However, different than the gradient boosting and random forest that have been considered in the supporting publication [169], the GPR model is exemplarily utilized here. Furthermore, several possibilities of considering time-varying operating conditions are investigated involving the operating condition information and the monitored condition-related signals. During the model development, the leave-one-out method is employed, such that sixteen out of the seventeen experimental data are employed for training, while the seventeenth previously unseen data is employed for testing.

Approach B: Condition-invariant features as inputs to a machine learning regression model

The first possibility investigated is utilizing the condition-invariant features derived from the monitored vibration signals, which was introduced in Section 5.3.2, as inputs to the GPR model. The number of inputs not only influences the accuracy of a machine learning model, but also the computational time required for training. Furthermore, some machine learning models perform better with scaled data. Thus, although a detailed investigation of these aspects is beyond the scope of this work, the number of input features and data scaling or min-max normalization were considered during the model development. Specifically, the influence of scaling the target variable, namely the RUL is considered here. To this end, the default parameters of an exponential GPR model is utilized, while ten top-ranked monotonic features, mostly from the HCTSA package are selected as input data and scaled to the $[0, 1]$ interval, during the investigations. The estimated RUL under consideration of the influence of scaling the RUL during training is shown in Figure 5.29(a). Both sub-figures of Figure 5.29(a) can generally be categorized in three segments. The first segment till about 15 min corresponds to the warm-up phase. The second segment from about 15 min till about 4.1 h corresponds to the healthy state after the warm-up phase and before the transition time, while the third segment is characterized by the faulty state. As presented in the top figure of Figure 5.29(a), although the RUL is significantly overestimated in the healthy state region, it declines towards the true RUL in the first segment, it possesses a constant trend with a large variation in the second segment and tends to the true RUL with a small variation in the third segment. Contrarily, as shown in the bottom figure of Figure 5.29(a), although the scaled estimated RUL is close to the true value in the first segment, it significantly fluctuates at an average in the second segment, thereafter it fluctuates with a significant deviation from the true RUL. Where the information of the true time at EoL is unknown, which is typical in online testing scenario, the estimated min-max normalized RUL cannot be easily transformed to the original units, typically the time or cycles remaining before the EoL. Thus, although the min-max normalization of the target RUL provides information about the degree of healthiness of a system, it lacks the information of the time or cycles remaining for maintenance decision making. Furthermore, especially for the experiment B02 exemplarily considered, it does not converge to the true RUL towards the end EoL, as can be seen in the bottom figure of Figure 5.29(a). Therefore, the target RUL is not min-max normalized in the following. To investigate the influence of considering more information during model training, the

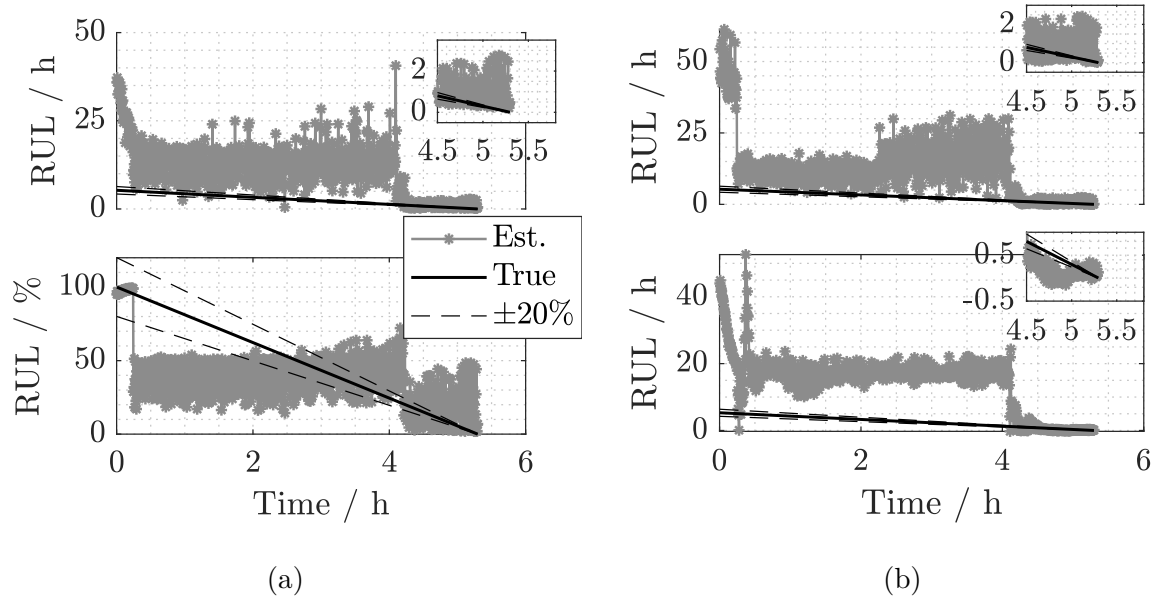


Figure 5.29: Estimated RUL for experiment B02 based on an exponential GPR model, under consideration of (a) the influence of target scaling and (b) top: twenty condition-invariant features as input; bottom: temperature data as further input

number of input features is doubled and the temperature information is also considered. As shown in the top figure of Figure 5.29(b), doubling the number of input features did not yield significant improvements for the considered experiment B02. Although it cannot be generalized, including redundant and non-trending features typically require more training time and often diminishes model accuracy. Thus, the top ten monotonic features are further considered. In some experiments, the temperature signal shows some significant increasing trend towards the EoL, thus the mean relative temperature T_2 is considered as a further input to the exponential GPR model. The resulting estimated RUL is shown in the bottom figure of Figure 5.29(b). In comparison to the estimation without the temperature information, the fluctuation is minimized and the estimated RUL is closer to the true RUL towards the EoL, as can be inferred from the bottom figure of Figure 5.29(b).

Approach C: Condition-centered data or scaled features as inputs to a machine learning regression model

Since the operating conditions, namely the rotating speed and bearing loads are simultaneously measured alongside the vibration signals, the features derived from the vibration signals can be scaled by the operating conditions. However, as discovered in Section 5.3.3, the speed-scaled vibration RMS has a more significant effect than the load-scaled vibration RMS. Furthermore, since the load-scaling could have an adverse effect on the resulting feature, as also discussed in Section 5.3.3, only the speed-scaling is considered here. Specifically, ten speed-scaled top-ranked classical time-domain features, including the vibration RMS, are utilized as inputs to train an exponential GPR model. For comparison, ten speed-scaled top-ranked condition-invariant features, mostly from the HCTSA package are also employed as inputs to train an exponential GPR model. The resulting estimated RUL for experiment B02 is exemplarily depicted in Figures 5.30(a) and

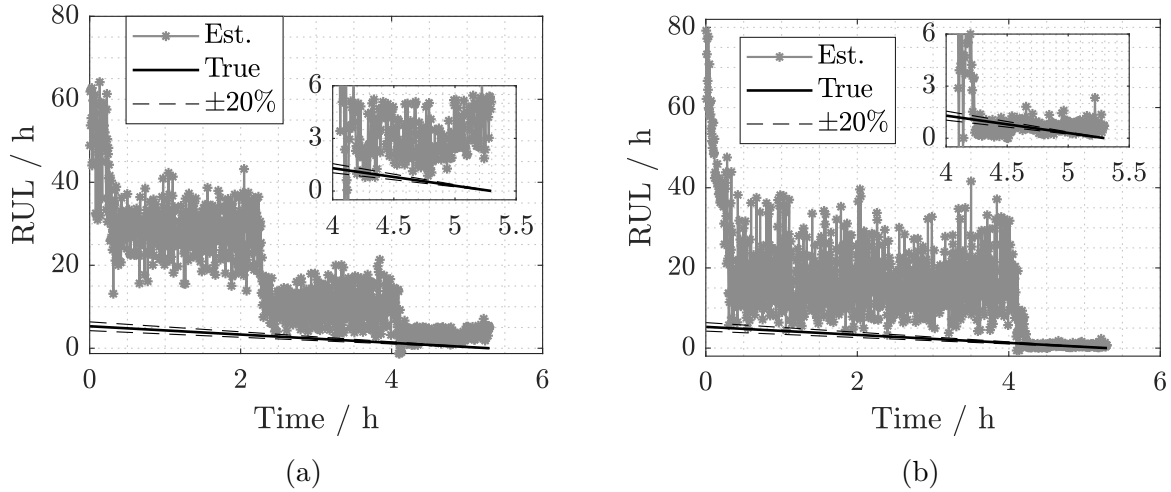


Figure 5.30: Estimated RUL for experiment B02 based on an exponential GPR model with ten speed-scaled inputs, namely (a) top-ranked classical time-domain features and (b) top-ranked condition-invariant features

5.30(b), under consideration of the speed-scaled classical features and the speed-scaled condition-invariant features, respectively. As shown in Figure 5.30(a), the estimated RUL lies approximately between $[30, 60]$ h in the warm-up phase. Although the true EoL is about five hours, the overestimation during the warm-up phase is realistic, owing to the relatively low operating conditions compared to operating conditions after the warm-up phase. The second portion of the figure, spanning from about 15 min to about 4 h, is characterized by a constant trend with large deviation from the true RUL. Furthermore, there is a slight decline in the estimated RUL at about 2.3 h and onward suggesting an increased degradation. However, in the faulty state, namely, at about 4.1 h and onward, the RUL is mostly overestimated and does not converge to the true RUL. In contrast, the estimations with the speed-scaled condition-invariant features tend to be closer to the true RUL in the faulty state. However, in the second portion of the figure, that is, from about 15 min to about 4 h, the estimation possesses an almost constant trend with a large variation.

Approach E: Embedding the operating conditions as further inputs to a machine learning regression model

The focus here is to investigate if incorporating the operating condition as further inputs to a machine learning regression model generally leads to a better model performance. To this end, the ten top-ranked condition-invariant features, are employed as inputs to train an exponential GPR model in addition to the operating condition information. The case where all available operating condition information is considered is compared against the case where only a significant operating condition is considered. Specifically, in the first case, the speed, static pre-load and dynamic load information are all considered as additional inputs to the model, similar to the already published work [169]. In the second case, only the speed information is considered as further information, because it was found to be more significant in Section 5.3.5. In both cases, all input data are min-max normalized, that is, scaled in the range $[0, 1]$, while the target is not min-max normalized. The resulting estimated RUL exemplarily for experiment B02 is shown in Figure 5.31(a), while considering all available operating condition information and in Figure 5.31(b), while

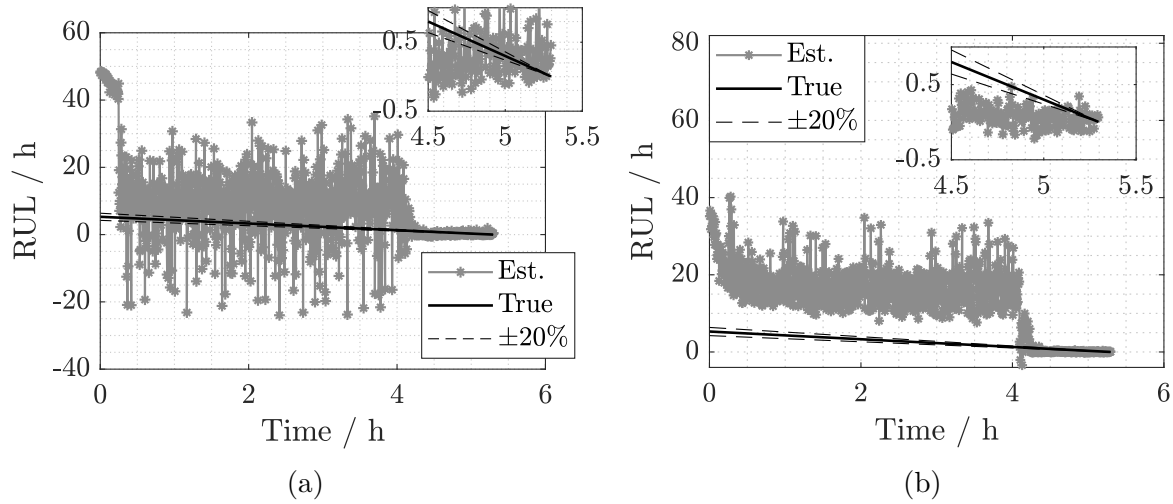


Figure 5.31: Estimated RUL for experiment B02 based on an exponential GPR model with ten condition-invariant features inputs in addition to (a) all available operating condition information and (b) only speed information

considering only the speed information. Both figures can also be partitioned into three segments, as previously described. The difference between both cases is more pronounced in the second portion, spanning from about 15 min to about 4 h, where incorporating all operating condition information introduces a large variation, as opposed to the case where only the speed information is considered. The large variation in Figure 5.31(a) is attributed to the nature of the dynamic load, which also possesses a large variation. Although the estimations from both models are close to the true RUL in the third region, that is, in the faulty state, the model with only the speed information is slightly better, because the model is more conservative, with lower variation.

5.5.3 Ensemble method

On the one hand, the modified bearing rating life did not incorporate the vibration features, but provided a theoretical RUL estimation close to the true RUL in the healthy segment, for some experiments. However, the RUL estimates fail to converge towards the bearing's EoL. On the other hand, the machine learning regression model incorporated the vibration features as inputs. The RUL estimation with the model was mostly significantly overestimated in the healthy segment, but close to the true RUL towards the bearing's EoL, because of the typically long non-trending healthy segment and short trending faulty segment of the features. Thus, an ensemble method that utilizes the modified bearing rating life to estimate the RUL in the healthy segment, but utilizes the exponential GPR model to estimate the RUL in the faulty segment, as depicted in Figure 5.32, is proposed here. The binary decision, if the system is in an healthy state or not, can be made via an expert-tailored threshold, or via machine learning methods, such as a binary clustering or classification method, as described in section 5.3. The transition point between healthy and faulty instance as determined by the clustering method with a condition-invariant feature, is exemplarily considered here. Furthermore, although different combinations of both models are conceivable, the model with the modified rating life under consideration of the temperature data is combined with the exponential GPR model with ten condition-

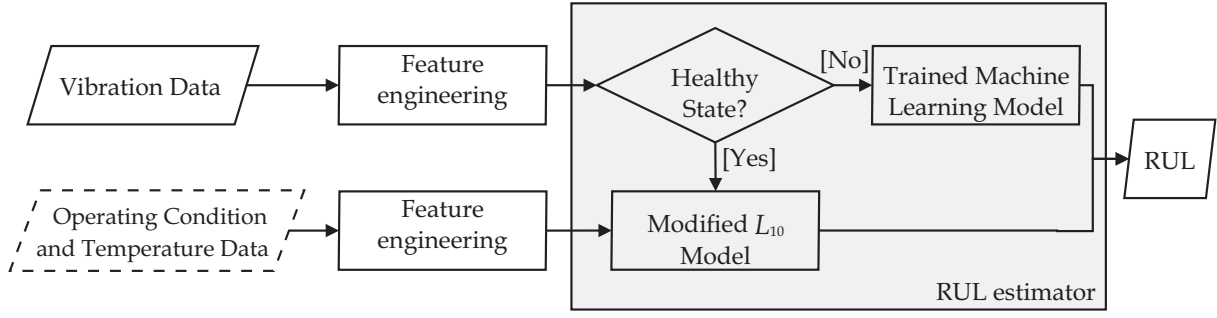


Figure 5.32: Ensemble method for estimating the RUL, considering non-trending healthy segment and short trending faulty segment of the vibration features

invariant features inputs in addition to the speed information. Specifically, on the one hand, the exponential GPR model is utilized to estimate the RUL after the transition time at about 4.13 h, as exemplarily seen in Figure 5.33(a). On the other hand, the 30th percentile of both the modified rating life and the exponential GPR model is exemplarily considered after the transition time to estimate the RUL, as exemplarily seen in Figure 5.33(b).

As can be deduced from both figures, the model performance is inherited from the corresponding model. Before the transition time, the model performance of the modified rating life is inherited in particular. However, after the transition time, the performance of the exponential GPR model is inherited, on the one hand, such that the estimation is heavily overestimated shortly after the transition time, but converges towards the true RUL, as can be inferred from Figure 5.33(a). On the other hand, the performance of both methods are regarded, such that the large variation of the modified rating life is generally reduced by the exponential GPR model, as can be observed from Figure 5.33(b). Furthermore, the overestimation of the exponential GPR model shortly after the transition time is also counteracted.

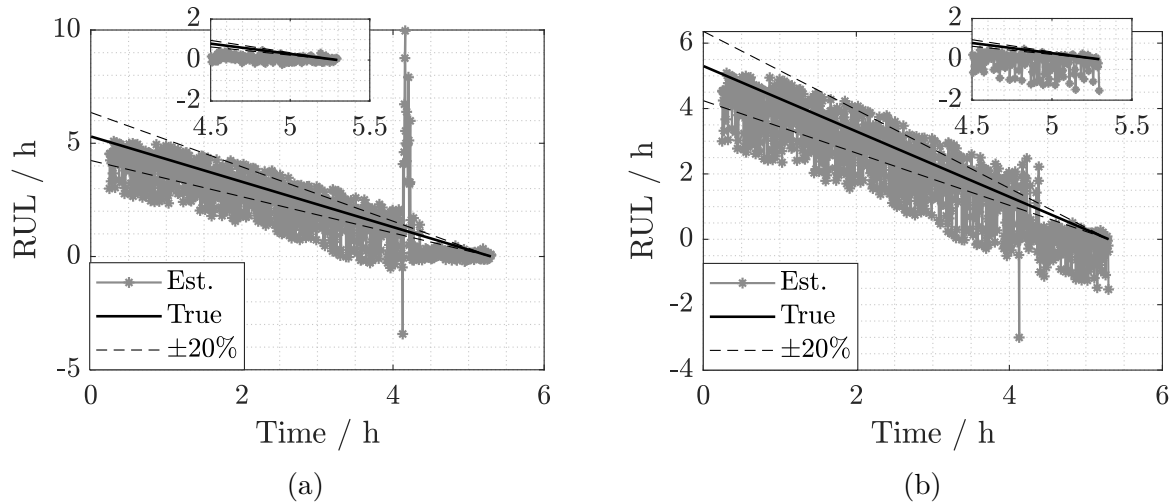


Figure 5.33: Estimated RUL for experiment B02 based on an ensemble model comprising of the modified rating life and (a) an exponential GPR model with ten condition-invariant features inputs in addition to the temperature information (b) an exponential GPR model with ten condition-invariant features inputs in addition to the speed information

5.6 Performance evaluation of the prognostics approaches

As previously stated, the ground truth discrete health state or continuous health index is inferred from the measured condition monitoring signals. Furthermore, since new bearings are installed during the run-to-failure experiments, the health index at the BoL assumes the value 1 and 0 at the EoL. As previously stated, the degradation is assumed to linearly accumulate. Thus, with time, the RUL is assumed to descend linearly from the BoL till the EoL. In this regard, a pseudo ground truth RUL with values in the $[0, 1]$ interval is derived as a basis for comparing the prognostics approaches. The RUL estimated by the considered models does not lie within the specified accuracy cone bounded at 20%, over the entire life time of the bearings and particularly did not remain within the bounds towards the EoL of the bearings. This results from the large variations and consequently significant uncertainty inherent of time-varying operations. For brevity, the comparative results based on the wMAPE is considered here for selected approaches, in the following. However, in the Appendix, Table 7 presents the evaluation results based on the nRMSE, while Table 6 presents the extended evaluation results based on the wMAPE.

In Table 5.4, approach LT considers the bearing rating life model with temperature-dependent modification factor. Approaches B, C and E utilize the exponential GPR model with ten condition-invariant features, while considering different levels of operating condition information. Approach B does not consider any further information, approach C considers speed-scaled features, while approach E considers the speed information, in addition to the condition invariant features. Furthermore, the best performing approach per bearing is highlighted via bold text. As presented in Table 5.4, approach LT outperforms the other presented approaches for most of the bearings, owing to the linear decrease of the RUL towards the end of life and the incorporation of the temperature-

dependent modification factor. However, for bearings B06, B07, B10 and B12, approach B outperforms the other approaches. Furthermore, approach B provides similar results to approach LT, for bearings B09 and B15. Approach C and approach E both lead to better performance than approach B for bearings B01, B05, B08, B13 and B16. However, approach E outperforms the other approaches only for bearings B01 and B08, while approach C outperforms the other approaches only for bearing B17.

In summary, approach LT, that is, the bearing rating life model with temperature-dependent modification factor, outperformed the other presented approaches, in most of the bearings. The next-best approach towards estimating the RUL of the considered ball bearings is the exponential GPR model trained with condition-invariant features engineered from the vibration signals (approach B). Although underperforming than the approach LT, scaling the features with respect to the speed (approach C) and the incorporation of the speed information as additional input to a machine learning model (approach E) provide some benefit towards estimating the RUL of ball bearings under time-varying operating conditions, particularly considering machine learning regression models.

Table 5.4: Performance evaluation with the wMAPE metric exemplarily for selected approaches

Bearing	Approach LT	Approach B	Approach C	Approach E
B01	1.54	4.05	3.96	1.28
B02	0.28	2.78	3.92	3.77
B03	0.33	11.57	9.91	11.85
B04	0.87	2.71	4.93	2.84
B05	1.05	15.05	9.06	8.31
B06	1.61	0.69	0.80	0.82
B07	1.59	0.53	0.84	0.83
B08	1.40	1.76	1.20	0.94
B09	1.04	1.04	1.29	1.33
B10	1.82	0.80	0.94	0.88
B11	0.33	0.54	1.61	0.79
B12	5.00	1.98	3.10	3.21
B13	0.56	0.84	0.68	0.68
B14	0.57	1.04	0.94	1.39
B15	0.64	0.64	0.68	0.71
B16	1.08	11.05	2.53	2.05
B17	4.53	2.57	1.74	2.58

6 Conclusion and Outlook

This concluding chapter will review the key research findings and examine their benefits and limitations, thereby laying the groundwork for future research opportunities.

Mechatronic and cyber-physical systems are subject to various changing environmental conditions and operate under time-varying operating conditions to fulfill specific goals. However, these time-varying conditions often have a negative impact on the degradation process and obscure condition monitoring signals, making it difficult to detect potential faults. This can result in significant economic losses and severe consequences for human safety and the environment, particularly in safety-critical systems. To address this challenge, three research objectives have been established to develop a comprehensive framework for reliable fault detection and prognostics in technical systems subjected to time-varying operating conditions.

The first objective was to investigate the interrelationship between the degradation process, the underlying operating conditions, and the condition monitoring signals. This study contributes to existing knowledge by demonstrating that while time-varying operating conditions, such as set by the operator or usage patterns, have a significant impact on both the degradation rate and the measured condition monitoring signals, these are not the sole contributing factors. The findings in this study underscore that other sources of variability, particularly those arising from material properties and manufacturing processes, also play a considerable role in influencing the degradation behavior. The interplay between these factors highlights the complexity of accurately detecting faults and predicting the RUL of technical systems.

The second objective aimed to identify features extracted from condition monitoring data that are more sensitive to degradation itself, rather than to variations in operating conditions. In practical terms, this means discovering features that reliably indicate fault progression, irrespective of (abrupt) jumps resulting from time-varying operating conditions. Through the application of advanced time-series analysis toolboxes, including HCTSA, EMG, and EEG feature extraction toolboxes, the study identifies a set of features that predominantly reflect the underlying degradation, even under changes in the operating conditions. These condition-invariant features provide a robust basis for fault detection and prognostics, even in the presence of time-varying operating conditions.

The third objective explored the methodologies and potential benefits associated with incorporating the operating conditions into fault detection and prognostics tasks. To address this, the research presents a comprehensive flowchart that systematizes five distinct approaches for leveraging information about time-varying operating conditions:

- A) **Specialized signal analysis techniques:** Adapting signal processing techniques to isolate or compensate for effects of specific operating conditions.
- B) **Condition-invariant feature engineering:** Extracting and selecting features that capture degradation independently of the underlying operating condition.
- C) **Condition-centered data or feature scaling:** Normalizing or scaling condition monitoring data or features derived thereof based on the current operating condition

to factor out its influence.

- D) **Condition-specific feature standardization and segmentation:** Segmenting and standardizing feature sets within distinct operating conditions to compensate the effects of varying operating conditions.
- E) **Embedding operating conditions in the modeling phase:** Directly including operating parameters as inputs in the model development phase, for example as a further input in machine learning models, thereby improving predictive accuracy.

By systematically integrating these approaches, the study demonstrates how models can more effectively account for the challenges posed by time-varying operating conditions, ultimately leading to reliable fault detection and prognostics.

In summary, the research not only highlights the diverse factors that affect system degradation under time-varying conditions, but also provides practical pathways for extracting meaningful degradation features and building robust fault detection and prognostics models. These contributions collectively advance the field of condition-based and predictive maintenance, especially for systems operating under time-varying conditions.

Although this study has provided practical pathways for addressing time-varying operating conditions, the following research opportunities have been identified to further enhance the accuracy of fault detection and prognostics models.

Firstly, in developing models to estimate the RUL of technical systems, where obtaining the true ground truth is challenging, the current state of the art often relies on a pseudo ground truth RUL based on assumptions from degradation models, such as a (piecewise) linear model. To improve performance evaluation, particularly for safety-critical systems, a promising research direction involves developing experimental procedures to validate the pseudo ground truth RUL or to leverage knowledge transfer from systems where the ground truth RUL can be directly measured.

Secondly, to better account for previously unseen operating conditions that can significantly affect the accuracy of fault detection and prognostics models, a valuable research direction is to extend the current study by incorporating a prediction of future time-varying operating conditions, along with an assessment of the associated prediction uncertainty.

Lastly, the current study focused on a limited failure mode observed under controlled experimental conditions. However, in real-world applications, certain fault or failure modes may not occur in some systems, while multiple fault or failure modes and mechanisms may occur simultaneously and interact with one another, particularly under time-varying operating conditions. Therefore, a key research direction is to extend the current study to address data imbalance issues, where fault or failure conditions represent the minority class. Additionally, further research should explore the use of transfer learning or on-line learning approaches to transfer or adapt existing models for handling multiple, new, and previously unseen fault or failure modes, as well as for application to other related technical systems.

References

- [1] R. Isermann and P. Ballé, “Trends in the application of model-based fault detection and diagnosis of technical processes”, *Control Engineering Practice*, vol. 5, no. 5, pp. 709–719, 1997.
- [2] H. M. Elattar, H. K. Elminir, and A. M. Riad, “Prognostics: A literature review”, *Complex & Intelligent Systems*, vol. 2, no. 2, pp. 125–154, 2016.
- [3] R. Al-Ali, L. Bulej, J. Kofroň, and T. Bureš, “A guide to design uncertainty-aware self-adaptive components in Cyber-Physical Systems”, *Future Generation Computer Systems*, vol. 128, pp. 466–489, 2022.
- [4] DIN 1311-1:2000-02(de), *(Mechanical) vibrations, oscillation and vibration systems - Part 1: Basic concepts, survey*. Deutsches Institut für Normung e.V. (DIN), Berlin, 2000.
- [5] VDI 4550 Part 1:2017-04, *Vibration analysis - Procedures and presentation of results - Terminology and basic principles*. Verein Deutscher Ingenieure e.V. (VDI), Düsseldorf, 2017.
- [6] B. Valeti and S. N. Pakzad, “Remaining useful life estimation of wind turbine blades under variable wind speed conditions using particle filters”, in *Proceedings of the Annual Conference of the PHM Society*, vol. 10, 2018.
- [7] I. de Pater and M. Mitici, “Developing health indicators and RUL prognostics for systems with few failure instances and varying operating conditions using a LSTM autoencoder”, *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105 582, 2023.
- [8] X. Xu, S. Tang, C. Yu, J. Xie, X. Han, and M. Ouyang, “Remaining useful life prediction of lithium-ion batteries based on wiener process under time-varying temperature condition”, *Reliability Engineering & System Safety*, vol. 214, p. 107 675, 2021.
- [9] J. Long, C. Chen, Z. Liu, J. Guo, and W. Chen, “Stochastic hybrid system approach to task-orientated remaining useful life prediction under time-varying operating conditions”, *Reliability Engineering & System Safety*, vol. 225, p. 108 568, 2022.
- [10] L. Bian, N. Gebraeel, and J. P. Kharoufeh, “Degradation modeling for real-time estimation of residual lifetimes in dynamic environments”, *IEEE Transactions*, vol. 47, no. 5, pp. 471–486, 2015.
- [11] J. Sim, J. Min, S. Kim, S. W. Lee, and J.-H. Choi, “Construction of bearing health indicator under time-varying operating conditions based on Isolation Forest”, *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107 058, 2023.
- [12] L. A. Kurgan and P. Musilek, “A survey of Knowledge Discovery and Data Mining process models”, *The Knowledge Engineering Review*, vol. 21, no. 1, pp. 1–24, 2006.
- [13] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases”, *AI magazine*, vol. 17, no. 3, pp. 37–37, 1996.

- [14] C. Shearer, “The CRISP-DM: the new blueprint for data mining”, *Journal of data warehousing*, vol. 5, no. 4, 2000.
- [15] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, “DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model”, *Procedia CIRP*, vol. 79, pp. 403–408, 2019.
- [16] G. Mariscal, O. Marban, and C. Fernandez, “A survey of data mining and knowledge discovery process models and methodologies”, *The Knowledge Engineering Review*, vol. 25, no. 2, pp. 137–166, 2010.
- [17] F. Martínez-Plumed, L. Contreras-Ochando, C. Ferri, J. Hernández-Orallo, M. Kull, et al., “CRISP-DM twenty years later: From data mining processes to data science trajectories”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048–3061, 2019.
- [18] ISO 13374-1:2003(en), *Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 1: General guidelines*. International Organization for Standardization (ISO), Geneva, Switzerland, 2003.
- [19] N. S. Clements, “Introduction to Prognostics”, in *Tutorial, Proceedings of the Annual Conference of the Prognostics and Health Management Society*, 2011.
- [20] ISO 13374-2:2007(en), *Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 2: Data processing*. International Organization for Standardization (ISO), Geneva, Switzerland, 2007.
- [21] ISO 13374-3:2012(en), *Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 3: Communication*. International Organization for Standardization (ISO), Geneva, Switzerland, 2012.
- [22] ISO 13374-4:2015(en), *Condition monitoring and diagnostics of machine systems - Data processing, communication and presentation - Part 4: Presentation*. International Organization for Standardization (ISO), Geneva, Switzerland, 2015.
- [23] MIMOSA, *Open System Architecture for Condition-Based Maintenance*, <https://www.mimosa.org/mimosa-osa-cbm/>, Accessed: 23.01.2024, 2024.
- [24] N. Omri, Z. Al Masry, N. Mairot, S. Giampiccolo, and N. Zerhouni, “Industrial data management strategy towards an SME-oriented PHM”, *Journal of Manufacturing Systems*, vol. 56, pp. 23–36, 2020.
- [25] T. Tinga and R. Loendersloot, “Aligning PHM, SHM and CBM by understanding the physical system failure behaviour”, in *PHM society European conference*, vol. 2, 2014.
- [26] ISO 17359:2018(en), *Condition monitoring and diagnostics of machines - General guidelines*. International Organization for Standardization (ISO), Geneva, Switzerland, 2018.
- [27] D. Li, Y. Wang, J. Wang, C. Wang, and Y. Duan, “Recent advances in sensor fault diagnosis: A review”, *Sensors and Actuators A: Physical*, vol. 309, p. 111 990, 2020.
- [28] P. M. Frank, “Evaluation of Analytical Redundancy for Fault Diagnosis in Dynamic Systems”, *IFAC Proceedings Volumes*, vol. 22, no. 6, pp. 7–19, 1989, IFAC Symposium on Advanced Information Processing in Automatic Control (AIPAC’89), Nancy, France, 3-5 July 1989.

-
- [29] A. Smiti, “A critical overview of outlier detection methods”, *Computer Science Review*, vol. 38, p. 100 306, 2020.
 - [30] S. Oberst, J. C. Lai, and T. A. Evans, “Physical basis of vibrational behaviour: Channel properties, noise and excitation signal extraction”, *Biotremology: Studying Vibrational Behavior*, pp. 53–78, 2019.
 - [31] P. Sevilla-Camacho, J. Robles-Ocampo, J. Muñoz-Soria, and F. Lee-Orantes, “Tool failure detection method for high-speed milling using vibration signal and reconfigurable bandpass digital filtering”, *The International Journal of Advanced Manufacturing Technology*, vol. 81, pp. 1187–1194, 2015.
 - [32] Y. Kopsinis and S. McLaughlin, “Development of EMD-based denoising methods inspired by wavelet thresholding”, *IEEE Transactions on signal Processing*, vol. 57, no. 4, pp. 1351–1362, 2009.
 - [33] L. van der Maaten, E. O. Postma, H. J. van den Herik, et al., “Dimensionality Reduction: A Comparative Review”, *Journal of Machine Learning Research*, vol. 10, no. 66-71, p. 13, 2009.
 - [34] B. Tang, W. Liu, and T. Song, “Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution”, *Renewable Energy*, vol. 35, no. 12, pp. 2862–2866, 2010.
 - [35] A. Bakiya, K. Kamalanand, V. Rajinikanth, R. S. Nayak, and S. Kadry, “Deep neural network assisted diagnosis of time-frequency transformed electromyograms”, *Multimedia Tools and Applications*, vol. 79, pp. 11 051–11 067, 2020.
 - [36] B. D. Fulcher, M. A. Little, and N. S. Jones, “Highly comparative time-series analysis: The empirical structure of time series and their methods”, *Journal of the Royal Society Interface*, vol. 10, no. 83, 2013.
 - [37] B. D. Fulcher and N. S. Jones, “hctsa: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction”, *Cell systems*, vol. 5, no. 5, pp. 527–531, 2017.
 - [38] J. Too, A. R. Abdullah, and N. M. Saad, “Classification of hand movements based on discrete wavelet transform and enhanced feature extraction”, *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 83–89, 2019.
 - [39] J. Too, A. R. Abdullah, N. Mohd Saad, and W. Tee, “EMG feature selection and classification using a Pbest-guide binary particle swarm optimization”, *Computation*, vol. 7, no. 1, p. 12, 2019.
 - [40] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, “Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)”, *Neurocomputing*, vol. 307, pp. 72–77, 2018.
 - [41] J. van der Donckt, J. van der Donckt, E. Deprost, and S. van Hoecke, “tsflex: Flexible time series processing & feature extraction”, *SoftwareX*, vol. 17, p. 100 971, 2022.
 - [42] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection”, *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
 - [43] R. Isermann, “Supervision, fault-detection and fault-diagnosis methods – An introduction”, *Control Engineering Practice*, vol. 5, no. 5, pp. 639–652, 1997.

- [44] M. Ulieru, “Diagnosis by approximate reasoning on dynamic fuzzy fault trees”, in *Proceedings of 1994 IEEE 3rd International Fuzzy Systems Conference*, 1994, pp. 2051–2056.
- [45] T. Höfling and T. Pfeufer, “Detection of additive and multiplicative faults-parity space vs. parameter estimation”, *IFAC Proceedings Volumes*, vol. 27, no. 5, pp. 515–520, 1994.
- [46] G. Fajdiga and M. Sraml, “Fatigue crack initiation and propagation under cyclic contact loading”, *Engineering fracture mechanics*, vol. 76, no. 9, pp. 1320–1335, 2009.
- [47] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, “A review of process fault detection and diagnosis: Part I: Quantitative model-based methods”, *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.
- [48] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, “A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies”, *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 313–326, 2003.
- [49] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, “A review of process fault detection and diagnosis: Part III: Process history based methods”, *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [50] S. Katipamula and M. R. Brambley, “Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems – A Review, Part I”, *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, 2005.
- [51] Y. Zhang and J. Jiang, “Bibliographical review on reconfigurable fault-tolerant control systems”, *Annual Reviews in Control*, vol. 32, no. 2, pp. 229–252, 2008.
- [52] A. Mouzakitidis, “Classification of fault diagnosis methods for control systems”, *Measurement and Control*, vol. 46, no. 10, pp. 303–308, 2013.
- [53] R. Arunthavanathan, F. Khan, S. Ahmed, and S. Imtiaz, “An analysis of process fault diagnosis methods from safety perspectives”, *Computers & Chemical Engineering*, vol. 145, p. 107197, 2021.
- [54] W. Quaghebeur, E. Torfs, B. De Baets, and I. Nopens, “Hybrid differential equations: Integrating mechanistic and data-driven techniques for modelling of water systems”, *Water Research*, vol. 213, p. 118166, 2022.
- [55] P. M. Frank, “Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results”, *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [56] R. Isermann, “Model-based fault-detection and diagnosis—status and applications”, *Annual Reviews in Control*, vol. 29, no. 1, pp. 71–85, 2005.
- [57] R. K. Mehra and J. Peschon, “An innovations approach to fault detection and diagnosis in dynamic systems”, *Automatica*, vol. 7, no. 5, pp. 637–640, 1971.
- [58] R. Isermann, “Supervision, fault-detection and fault-diagnosis methods – an introduction”, *Control Engineering Practice*, vol. 5, no. 5, pp. 639–652, 1997.
- [59] Y. Tsukamoto and T. Terano, “Failure diagnosis by using fuzzy logic”, in *1977 IEEE Conference on Decision and Control*, 1977, pp. 1390–1395.

-
- [60] S. M. Islam, T. Wu, and G. Ledwich, “A novel fuzzy logic approach to transformer fault diagnosis”, *IEEE Transactions on Dielectrics and electrical Insulation*, vol. 7, no. 2, pp. 177–186, 2000.
 - [61] S. Gollapudi, *Practical Machine Learning*. Packt Publishing, 2016, ISBN: 97817843-94011.
 - [62] S. Few, “Quantitative vs. Categorical Data; A Difference Worth Knowing”, *Information Management*, vol. 15, no. 4, p. 42, 2005.
 - [63] D. An, N. H. Kim, and J.-H. Choi, “Practical options for selecting data-driven or physics-based prognostics algorithms with reviews”, *Reliability Engineering & System Safety*, vol. 133, pp. 223–236, 2015.
 - [64] K. Javed, R. Gouriveau, and N. Zerhouni, “State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels”, *Mechanical Systems and Signal Processing*, vol. 94, pp. 214–236, 2017.
 - [65] T. Sutharssan, D. Montalvao, Y. K. Chen, W.-C. Wang, C. Pisac, and H. Elemara, “A review on prognostics and health monitoring of proton exchange membrane fuel cell”, *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 440–450, 2017.
 - [66] J. Guo, Z. Li, and M. Li, “A review on prognostics methods for engineering systems”, *IEEE Transactions on Reliability*, vol. 69, no. 3, pp. 1110–1129, 2019.
 - [67] M. S. Ibrahim, J. Fan, W. K. Yung, A. Prisacaru, W. van Driel, X. Fan, and G. Zhang, “Machine learning and digital twin driven diagnostics and prognostics of light-emitting diodes”, *Laser & Photonics Reviews*, vol. 14, no. 12, 2020.
 - [68] Y. Peng, M. Dong, and M. J. Zuo, “Current status of machine prognostics in condition-based maintenance: A review”, *The International Journal of Advanced Manufacturing Technology*, vol. 50, pp. 297–313, 2010.
 - [69] J. J. M. Jimenez, S. Schwartz, R. Vingerhoeds, B. Grabot, and M. Salaün, “Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics”, *Journal of Manufacturing Systems*, vol. 56, pp. 539–557, 2020.
 - [70] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, “Rotating machinery prognostics: State of the art, challenges and opportunities”, *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.
 - [71] J. Gu and M. Pecht, “Prognostics and health management using physics-of-failure”, in *2008 Annual Reliability and Maintainability Symposium*, IEEE, 2008, pp. 481–487.
 - [72] M. Pecht and J. Gu, “Physics-of-failure-based prognostics for electronic products”, *Transactions of the Institute of Measurement and Control*, vol. 31, no. 3-4, pp. 309–322, 2009.
 - [73] O. F. Eker, F. Camci, and I. K. Jennions, “Major challenges in prognostics: Study on benchmarking prognostics datasets”, in *PHM Society European Conference*, vol. 1, 2012.
 - [74] F. Zhao, Z. Tian, and Y. Zeng, “Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method”, *IEEE Transactions on Reliability*, vol. 62, no. 1, pp. 146–159, 2013.

- [75] H. Oh, B. Han, P. McCluskey, C. Han, and B. D. Youn, “Physics-of-failure, condition monitoring, and prognostics of insulated gate bipolar transistor modules: A review”, *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2413–2426, 2014.
- [76] V. Atamuradov, K. Medjaher, P. Dersin, B. Lamoureux, and N. Zerhouni, “Prognostics and health management for maintenance practitioners-Review, implementation and tools evaluation.”, *International Journal of Prognostics and Health Management*, vol. 8, no. 3, pp. 1–31, 2017.
- [77] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, “A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery”, *Advances in Mechanical Engineering*, vol. 8, no. 8, 2016.
- [78] H. Tanizaki, “Prediction and Smoothing”, in *Nonlinear Filters: Estimation and Applications*. Springer Berlin Heidelberg, 1996, pp. 205–231, ISBN: 978-3-662-03223-7.
- [79] Z. Chen et al., “Bayesian filtering: From Kalman filters to particle filters, and beyond”, *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.
- [80] S. P. Won, W. W. Melek, and F. Golnaraghi, “A Kalman/Particle Filter-Based Position and Orientation Estimation Method Using a Position Sensor/Inertial Measurement Unit Hybrid System”, *IEEE Transactions on Industrial Electronics*, vol. 57, no. 5, pp. 1787–1798, 2009.
- [81] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni, “Particle filter-based prognostics: Review, discussion and perspectives”, *Mechanical Systems and Signal Processing*, vol. 72, pp. 2–31, 2016.
- [82] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, “Prognostic modelling options for remaining useful life estimation by industry”, *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1803–1836, 2011.
- [83] J. M. van Noortwijk, “A survey of the application of gamma processes in maintenance”, *Reliability Engineering & System Safety*, vol. 94, no. 1, pp. 2–21, 2009.
- [84] Z. Zhang, X. Si, C. Hu, and Y. Lei, “Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods”, *European Journal of Operational Research*, vol. 271, no. 3, pp. 775–796, 2018.
- [85] S. Tang, C. Yu, X. Wang, X. Guo, and X. Si, “Remaining Useful Life Prediction of Lithium-Ion Batteries Based on the Wiener Process with Measurement Error”, *Energies*, vol. 7, no. 2, pp. 520–547, 2014.
- [86] O. K. Aimiyeagbon, A. Bender, and W. Sextro, “Evaluation of time series forecasting approaches for the reliable crack length prediction of riveted aluminium plates given insufficient data”, in *Proceedings of the European Conference of the PHM Society*, vol. 5, 2020, pp. 1–11.
- [87] G. Wang and J. Xiang, “Remain useful life prediction of rolling bearings based on exponential model optimized by gradient method”, *Measurement*, vol. 176, p. 109 161, 2021.
- [88] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd edition. OTexts: Melbourne, Australia, 2021, Accessed: 26.02.2024. [Online]. Available: <https://otexts.com/fpp3/>

-
- [89] B. Long, W. Xian, L. Jiang, and Z. Liu, “An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries”, *Microelectronics Reliability*, vol. 53, no. 6, pp. 821–831, 2013.
 - [90] Y. Zhou and M. Huang, “Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model”, *Microelectronics Reliability*, vol. 65, pp. 265–273, 2016.
 - [91] M. Fernández-Delgado, M. S. Sirsat, E. Cernadas, S. Alawadi, S. Barro, and M. Febrero-Bande, “An extensive experimental survey of regression methods”, *Neural Networks*, vol. 111, pp. 11–34, 2019.
 - [92] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, “Linear regression”, in *An Introduction to Statistical Learning: with Applications in Python*, Springer, 2023, pp. 69–134.
 - [93] V. Vapnik, S. E. Golowich, and A. J. Smola, “Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing”, in *Advances in Neural Information Processing Systems*, 1997, pp. 281–287.
 - [94] V. N. Vapnik, *Statistical Learning Theory*. Wiley New York, 1998, vol. 1.
 - [95] B. Samanta, “Gear fault detection using artificial neural networks and support vector machines with genetic algorithms”, *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 625–644, 2004.
 - [96] N. Z. Gebraeel, M. Lawley, et al., “A neural network degradation model for computing and updating residual life distributions”, *Automation Science and Engineering, IEEE Transactions on*, vol. 5, no. 1, pp. 154–163, 2008.
 - [97] A. Kumar, A. Srivastava, A. Banerjee, and A. Goel, “Performance Based Anomaly Detection Analysis of a Gas Turbine Engine by Artificial Neural Network Approach”, in *Annual Conference of Prognostics and Health Management Society*, 2012.
 - [98] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
 - [99] P. P. Shinde and S. Shah, “A Review of Machine Learning and Deep Learning Applications”, in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, 2018, pp. 1–6.
 - [100] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, and M. Wei, “A Review on Deep Learning Applications in Prognostics and Health Management”, *IEEE Access*, vol. 7, pp. 162 415–162 438, 2019.
 - [101] A. Tharwat, “Classification assessment methods”, *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2021.
 - [102] C. Ferri, J. Hernández-Orallo, and R. Modroiu, “An experimental comparison of performance measures for classification”, *Pattern Recognition Letters*, vol. 30, no. 1, pp. 27–38, 2009.
 - [103] S. B. Johnson, S. Ghoshal, D. Haste, and C. E. Moore, “Fault Management Metrics”, in *AIAA Information Systems-AIAA Infotech@ Aerospace*, 2017, p. 1295.
 - [104] J. Lever, M. Krzywinski, and N. Altman, “Classification evaluation: It is important to understand both what a classification metric expresses and what it hides”, *Nature Methods*, vol. 13, no. 8, pp. 603–605, 2016.

- [105] V. García, R. A. Mollineda, and J. S. Sánchez, “Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions”, in *Iberian Conference on Pattern Recognition and Image Analysis*, Springer, 2009, pp. 441–448.
- [106] X. Cheng, F. Shi, X. Liu, M. Zhao, and S. Chen, “A Novel Deep Class-Imbalanced Semisupervised Model for Wind Turbine Blade Icing Detection”, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2558–2570, 2021.
- [107] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, “Metrics for evaluating performance of prognostic techniques”, in *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1–17.
- [108] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, “Metrics for Offline Evaluation of Prognostic Performance”, *International Journal of Prognostics and health management*, vol. 1, no. 1, pp. 4–23, 2010.
- [109] A. Botchkarev, “A new typology design of performance metrics to measure errors in machine learning regression algorithms”, *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, pp. 045–076, 2019.
- [110] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy”, *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
- [111] S. Makridakis, “Accuracy measures: Theoretical and practical concerns”, *International journal of forecasting*, vol. 9, no. 4, pp. 527–529, 1993.
- [112] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, “On applying the prognostic performance metrics”, in *Annual Conference of the PHM Society*, vol. 1, 2009.
- [113] M. Naser and A. H. Alavi, “Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences”, *Architecture, Structures and Construction*, vol. 3, no. 4, pp. 499–517, 2023.
- [114] S. Ochella and M. Shafiee, “Performance metrics for artificial intelligence (AI) algorithms adopted in prognostics and health management (PHM) of mechanical systems”, in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1828, 2021, p. 012005.
- [115] S. Kolassa and W. Schütz, “Advantages of the MAD/MEAN ratio over the MAPE”, *Foresight: The International Journal of Applied Forecasting*, no. 6, pp. 40–43, 2007.
- [116] D. Butler, “The shock-pulse method for the detection of damaged rolling bearings”, *Non-Destructive Testing*, vol. 6, no. 2, pp. 92–95, 1973.
- [117] ISO 18431-4:2007(en), *Mechanical vibration and shock - Signal processing - Part 4: Shock-response spectrum analysis*. International Organization for Standardization (ISO), Geneva, Switzerland, 2007.
- [118] R. Zhang, J. Kang, L. Hao, X. Zhang, H. Teng, and H. Li, “Research on operating condition effect on the shock pulse method”, *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, no. 7, pp. 5392–5398, 2014.
- [119] Y. He, M. Hu, K. Feng, and Z. Jiang, “Bearing condition evaluation based on the shock pulse method and principal resonance analysis”, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

-
- [120] D. R. Houser and M. J. Drosjack, *Vibration Signal Analysis Techniques*. Eustis Directorate, US Army Air Mobility and Development Laboratory, 1973.
 - [121] K. Bossley, R. Mckendrick, C. Harris, and C. Mercer, “Hybrid computed order tracking”, *Mechanical Systems and Signal Processing*, vol. 13, no. 4, pp. 627–641, 1999.
 - [122] K. Fyfe and E. Munck, “Analysis of computed order tracking”, *Mechanical Systems and Signal Processing*, vol. 11, no. 2, pp. 187–205, 1997.
 - [123] S. Lu, R. Yan, Y. Liu, and Q. Wang, “Tacholeless speed estimation in order tracking: A review with application to rotating machine fault diagnosis”, *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 7, pp. 2315–2332, 2019.
 - [124] C. Peeters, Q. Leclere, J. Antoni, P. Lindahl, J. Donnal, S. Leeb, and J. Helsen, “Review and comparison of tacholeless instantaneous speed estimation methods on experimental vibration data”, *Mechanical Systems and Signal Processing*, vol. 129, pp. 407–436, 2019.
 - [125] S. Liu, C. Liu, Y. Wu, S. Liu, and W. Liu, “Tacholeless Order Tracking Method Based on STFrFT for Fault Diagnosis of Wind Turbine Gearbox Under Variable Operating Conditions”, *IEEE Transactions on Instrumentation and Measurement*, 2024.
 - [126] I. Howard and I. M. Howard, “A Review of Rolling Element Bearing Vibration "Detection, Diagnosis and Prognosis"", *Aeronautical and Maritime Research Laboratory, Defence Science and Technology Organisation (DSTO)*, no. DSTO-RR-0013, 1994.
 - [127] Z. Peng, W. T. Peter, and F. Chu, “A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing”, *Mechanical Systems and Signal Processing*, vol. 19, no. 5, pp. 974–988, 2005.
 - [128] B. Hou, D. Wang, Y. Wang, T. Yan, Z. Peng, and K.-L. Tsui, “Adaptive weighted signal preprocessing technique for machine health monitoring”, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2020.
 - [129] B. Hou, D. Wang, T. Xia, Y. Wang, Y. Zhao, and K.-L. Tsui, “Investigations on quasi-arithmetic means for machine condition monitoring”, *Mechanical Systems and Signal Processing*, vol. 151, p. 107451, 2021.
 - [130] J. Singh, A. Darpe, and S. Singh, “Bearing damage assessment using Jensen–Rényi Divergence based on EEMD”, *Mechanical Systems and Signal Processing*, vol. 87, pp. 307–339, 2017.
 - [131] S. Schmidt, P. S. Heyns, and K. C. Gryllias, “An informative frequency band identification framework for gearbox fault diagnosis under time-varying operating conditions”, *Mechanical Systems and Signal Processing*, vol. 158, p. 107771, 2021.
 - [132] S. Schmidt, P. S. Heyns, and K. C. Gryllias, “Combining the Spectral Coherence with Informative Frequency Band Features for Condition Monitoring Under Time-Varying Operating Conditions”, in *International Congress and Workshop on Industrial AI*, Springer, 2021, pp. 189–201.

- [133] X. Wang, T. Wang, A. Ming, W. Zhang, A. Li, and F. Chu, “Cross-operating condition degradation knowledge learning for remaining useful life estimation of bearings”, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [134] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, “PRONOSTIA: An experimental platform for bearings accelerated degradation tests.”, in *IEEE International Conference on Prognostics and Health Management, PHM’12.*, IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.
- [135] K. Rombach, G. Michau, and O. Fink, “Contrastive learning for fault detection and diagnostics in the context of changing operating conditions and novel fault types”, *Sensors*, vol. 21, no. 10, p. 3550, 2021.
- [136] G. Michau and O. Fink, “Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer”, *Knowledge-Based Systems*, vol. 216, p. 106816, 2021.
- [137] O. K. Aimiyeagbon, A. Bender, and W. Sextro, “On the applicability of time series features as health indicators for technical systems operating under varying conditions”, in *Proceedings of the Seventeenth International Conference on Condition Monitoring and Asset Management (CM 2021)*, 2021.
- [138] R.-V. Sánchez, P. Lucero, J.-C. Macancela, M. Cerrada, R. E. Vásquez, and F. Pacheco, “Multi-fault Diagnosis of Rotating Machinery by Using Feature Ranking Methods and SVM-based Classifiers”, in *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, IEEE, 2017, pp. 105–110.
- [139] B. Nayana and P. Geethanjali, “Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal”, *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5618–5625, 2017.
- [140] J. Too. “EMG README”. Accessed: 04.2024. [Online]. Available: <https://github.com/JingweiToo/EMG-Feature-Extraction-Toolbox>
- [141] J. Bogatinovski, D. Kocev, and A. Rashkovska, “Feature Extraction for Heart-beat Classification in Single-Lead ECG”, in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2019, pp. 320–325.
- [142] C. Murphy-Royal, A. D. Johnston, A. K. Boyce, B. Diaz-Castro, A. Institoris, G. Peringod, O. Zhang, R. F. Stout, D. C. Spray, R. J. Thompson, et al., “Stress gates an astrocytic energy reservoir to impair synaptic plasticity”, *Nature Communications*, vol. 11, no. 1, pp. 1–18, 2020.
- [143] J. Coble and J. W. Hines, “Identifying optimal prognostic parameters from data: A genetic algorithms approach”, in *Annual Conference of the PHM Society*, vol. 1, 2009.
- [144] J. A. Carino, D. Zurita, M. Delgado, J. A. Ortega, and R. J. Romero-Troncoso, “Remaining useful life estimation of ball bearings by means of monotonic score calibration”, in *2015 IEEE International Conference on Industrial Technology (ICIT)*, 2015, pp. 1752–1758.

-
- [145] C. Stander, P. Heyns, and W. Schoombie, “Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions”, *Mechanical Systems and Signal Processing*, vol. 16, no. 6, pp. 1005–1024, 2002.
 - [146] S. Schmidt and P. S. Heyns, “Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring”, *Measurement*, vol. 149, p. 106964, 2020.
 - [147] T. Chrisp, G. Starrs, W. J. McCarter, E. Rouchotas, and J. Blewett, “Temperature-conductivity relationships for concrete: An activation energy approach”, *Journal of Materials Science Letters*, vol. 20, pp. 1085–1087, 2001.
 - [148] B. Jha and L. Dong, “Lifetime Improvement With Predictive Maintenance of Power Electronics Based on Remaining Useful Life Prediction”, in *2024 IEEE Texas Power and Energy Conference (TPEC)*, IEEE, 2024, pp. 1–6.
 - [149] M. Rigamonti, P. Baraldi, E. Zio, D. Astigarraga, and A. Galarza, “Particle filter-based prognostics for an electrolytic capacitor working in variable operating conditions”, *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1567–1575, 2015.
 - [150] A. J. Croxford, J. Moll, P. D. Wilcox, and J. E. Michaels, “Efficient temperature compensation strategies for guided wave structural health monitoring”, *Ultrasonics*, vol. 50, no. 4-5, pp. 517–528, 2010.
 - [151] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “Efficient backprop”, in *Neural networks: Tricks of the trade*, Springer, 2002, pp. 9–50.
 - [152] A. Saxena and K. Goebel, *PHM08 Challenge Data Set. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA*, 2008.
 - [153] L. Peel, “Data driven prognostics using a Kalman filter ensemble of neural network models”, in *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1–6.
 - [154] T. Wang, J. Yu, D. Siegel, and J. Lee, “A similarity-based prognostics approach for remaining useful life estimation of engineered systems”, in *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1–6.
 - [155] D. Xu, X. Xiao, J. Liu, and S. Sui, “Spatio-temporal degradation modeling and remaining useful life prediction under multiple operating conditions based on attention mechanism and deep learning”, *Reliability Engineering & System Safety*, vol. 229, p. 108886, 2023.
 - [156] A. Javanmardi and E. Hüllermeier, “Conformal Prediction Intervals for Remaining Useful Lifetime Estimation”, *International Journal of Prognostics and Health Management*, vol. 14, no. 2, 2023.
 - [157] M. L. Baptista, E. M. Henriques, and K. Goebel, “A self-organizing map and a normalizing multi-layer perceptron approach to baselining in prognostics under dynamic regimes”, *Neurocomputing*, vol. 456, pp. 268–287, 2021.
 - [158] Z. Zhang, X. Chen, and E. Zio, “A framework for predicting the remaining useful life of machinery working under time-varying operational conditions”, *Applied Soft Computing*, vol. 126, p. 109164, 2022.

- [159] S. Kim, H. J. Park, Y.-H. Seo, and J.-H. Choi, "A robust health indicator for rotating machinery under time-varying operating conditions", *IEEE Access*, vol. 10, pp. 4993–5001, 2022.
- [160] S. Roy, K. Lonkar, V. Janapati, and F.-K. Chang, "A novel physics-based temperature compensation model for structural health monitoring using ultrasonic guided waves", *Structural Health Monitoring*, vol. 13, no. 3, pp. 321–342, 2014.
- [161] H. Wang, H. Liao, and X. Ma, "Remaining useful life prediction considering joint dependency of degradation rate and variation on time-varying operating conditions", *IEEE Transactions on Reliability*, vol. 70, no. 2, pp. 761–774, 2020.
- [162] N. Li, N. Gebraeel, Y. Lei, L. Bian, and X. Si, "Remaining useful life prediction of machinery under time-varying operating conditions based on a two-factor state-space model", *Reliability Engineering & System Safety*, vol. 186, pp. 88–100, 2019.
- [163] Y. A. Yucesan and F. A. Viana, "A physics-informed neural network for wind turbine main bearing fatigue", *International Journal of Prognostics and Health Management*, vol. 11, no. 1, 2020.
- [164] K. Geurts, K. Eryilmaz, and T. Oijevaar, "A Sequential Hybrid Method for Full Lifetime Remaining Useful Life Prediction of Bearings in Rotating Machinery", in *Annual Conference of the PHM Society*, vol. 15, 2023.
- [165] N. Kayedpour, J. Qing, J. Wauters, J. D. De Kooning, I. Couckuyt, and G. Crevecoeur, "Wind Turbine Hybrid Physics-based Deep Learning Model for a Health Monitoring Approach Considering Provision of Ancillary Services", *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [166] J. Shi, A. Rivera, and D. Wu, "Battery health management using physics-informed machine learning: Online degradation modeling and remaining useful life prediction", *Mechanical Systems and Signal Processing*, vol. 179, p. 109347, 2022.
- [167] G. J. Bowden, G. C. Dandy, and H. R. Maier, "Input determination for neural network models in water resources applications. part 1—background and methodology", *Journal of Hydrology*, vol. 301, no. 1-4, pp. 75–92, 2005.
- [168] N. Muttill and K.-W. Chau, "Machine-learning paradigms for selecting ecologically significant input variables", *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 735–744, 2007.
- [169] A. Javanmardi, O. K. Aimiyekagbon, A. Bender, J. K. Kimotho, W. Sextro, and E. Hüllermeier, "Remaining Useful Lifetime Estimation of Bearings Operating under Time-Varying Conditions", in *PHM Society European Conference*, vol. 8, 2024.
- [170] K. Uchino, "Chapter 1 - The Development of Piezoelectric Materials and the New Perspective", in *Advanced Piezoelectric Materials*, Second Edition, Woodhead Publishing, 2017, pp. 1–92, ISBN: 978-0-08-102135-4.
- [171] E. K. Akdogan, M. Allahverdi, and A. Safari, "Piezoelectric Composites for Sensor and Actuator Applications", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 52, no. 5, pp. 746–775, 2005.
- [172] G. Park, C. R. Farrar, F. L. di Scalea, and S. Coccia, "Performance assessment and validation of piezoelectric active-sensors in structural health monitoring", *Smart Materials and Structures*, vol. 15, no. 6, p. 1673, 2006.

-
- [173] T. Hemsell, P. Bornmann, T. Morita, C. Sondermann-Wölke, and W. Sextro, “Reliability analysis of ultrasonic power transducers”, *Archive of Applied Mechanics*, vol. 86, pp. 1707–1713, 2016.
 - [174] J. K. Kimotho, T. Hemsell, and W. Sextro, “Estimation of remaining useful lifetime of piezoelectric transducers based on self-sensing”, *IEEE Transactions on Reliability*, vol. 66, no. 3, pp. 914–923, 2017.
 - [175] O. K. Aimiyeagbon, A. Bender, T. Hemsell, and W. Sextro, “Diagnostics of Piezoelectric Bending Actuators Subjected to Varying Operating Conditions”, *Electronics*, vol. 13, no. 3, 2024.
 - [176] M. Kuna, “Fracture mechanics of piezoelectric materials—where are we right now?”, *Engineering Fracture Mechanics*, vol. 77, no. 2, pp. 309–326, 2010.
 - [177] F. Lowrie, M. G. Cain, M. Stewart, and M. G. Gee, “Time dependent behaviour of piezo-electric materials.”, *National Physical Laboratory (NPL) Report*, no. CMMT(A)151, 1999.
 - [178] D. Van den Ende, B. Bos, and W. Groen, “Non-linear electromechanical behaviour of piezoelectric bimorph actuators: Influence on performance and lifetime”, *Journal of Electroceramics*, vol. 22, no. 1-3, pp. 185–191, 2009.
 - [179] R. Dunsch and J.-M. Breguet, “Unified mechanical approach to piezoelectric bender modeling”, *Sensors and Actuators A: physical*, vol. 134, no. 2, pp. 436–446, 2007.
 - [180] S. A. Rios and A. J. Fleming, “A new electrical configuration for improving the range of piezoelectric bimorph benders”, *Sensors and Actuators A: Physical*, vol. 224, pp. 106–110, 2015.
 - [181] A. Bender, “Model-based condition monitoring of piezoelectric bending actuators”, *Sensors and Actuators A: Physical*, vol. 357, p. 114 399, 2023.
 - [182] A. K. Tagantsev, I. Stolichnov, E. L. Colla, and N. Setter, “Polarization fatigue in ferroelectric films: Basic experimental findings, phenomenological scenarios, and microscopic features”, *Journal of Applied Physics*, vol. 90, no. 3, pp. 1387–1402, 2001.
 - [183] M. P. Denzler, *Lebensdauer und zuverlässigkeit dynamisch betriebener piezokeramischer Biegeumwandler*. Göttingen: Cuvillier Verlag Göttingen, 2004.
 - [184] K. Ruschmeyer, *Piezokeramik: Grundlagen, Werkstoffe, Applikationen* (Kontakt & Studium). Renningen-Malmsheim: Expert-Verl., 1995, vol. Bd. 460 : Werkstoffe, ISBN: 3816911528.
 - [185] E.-M. Anton, W. Jo, D. Damjanovic, and J. Rödel, “Determination of depolarization temperature of $(\text{Bi}_{1/2}\text{Na}_{1/2})\text{TiO}_3$ -based lead-free piezoceramics”, *Journal of Applied Physics*, vol. 110, no. 9, 2011.
 - [186] H.-C. Liu, L. Liu, and N. Liu, “Risk evaluation approaches in failure mode and effects analysis: A literature review”, *Expert Systems with Applications*, vol. 40, no. 2, pp. 828–838, 2013.
 - [187] V. Govorukha, M. Kamlah, V. Loboda, and Y. Lapusta, “Interface cracks in piezoelectric materials”, *Smart Materials and Structures*, vol. 25, no. 2, p. 023 001, 2016.
 - [188] M. Shoaib, N. H. Hamid, M. Tariq Jan, and N. B. Zain Ali, “Effects of Crack Faults on the Dynamics of Piezoelectric Cantilever-Based MEMS Sensor”, *IEEE Sensors Journal*, vol. 17, no. 19, pp. 6279–6294, 2017.

- [189] S. Zeng, C. Zhang, K. Wang, B. Wang, and L. Sun, “Analysis of delamination of unimorph cantilever piezoelectric energy harvesters”, *Journal of Intelligent Material Systems and Structures*, vol. 29, no. 9, pp. 1875–1883, 2018.
- [190] R. Salazar, M. Serrano, and A. Abdelkefi, “Fatigue in piezoelectric ceramic vibrational energy harvesting: A review”, *Applied Energy*, vol. 270, p. 115 161, 2020.
- [191] T. Hemsel, J. Krome, W. Littmann, and J. Wallaschek, “Modellierung piezoelektrischer Stapel- und Vielschichtaktoren für dynamischen Betrieb”, in *Polnisch-Deutscher Workshop (Werkzeuge der Mechatronik)*, 1998.
- [192] S.-F. Ling and Y. Xie, “Detecting mechanical impedance of structures using the sensing capability of a piezoceramic inertial actuator”, *Sensors and Actuators A: Physical*, vol. 93, no. 3, pp. 243–249, 2001.
- [193] W. Littmann, H. Storck, and J. Wallaschek, “Reduction of friction using piezoelectrically excited ultrasonic vibrations”, in *Smart Structures and Materials 2001: Damping and Isolation*, SPIE, vol. 4331, 2001, pp. 302–311.
- [194] J. Twiefel, B. Richter, T. Sattel, and J. Wallaschek, “Power output estimation and experimental validation for piezoelectric energy harvesting systems”, *Journal of Electroceramics*, vol. 20, pp. 203–208, 2008.
- [195] P. V. Avvari, Y. Yang, and C. K. Soh, “Long-term fatigue behavior of a cantilever piezoelectric energy harvester”, *Journal of Intelligent Material Systems and Structures*, vol. 28, no. 9, pp. 1188–1210, 2017.
- [196] L. Yao, J. Zhang, L. Lu, and M. Lai, “Nonlinear dynamic characteristics of piezoelectric bending actuators under strong applied electric field”, *Journal of Microelectromechanical Systems*, vol. 13, no. 4, pp. 645–652, 2004.
- [197] T. Li, Y. Chen, and J. Ma, “Frequency dependence of piezoelectric vibration velocity”, *Sensors and actuators A: physical*, vol. 138, no. 2, pp. 404–410, 2007.
- [198] W. Littmann, T. Hemsel, C. Kauczor, J. Wallaschek, and M. Sinha, “Load-adaptive phase-controller for resonant driven piezoelectric devices”, in *World Congress Ultrasonics, Paris*, vol. 48, 2003, p. 64.
- [199] J. Hughes and P. G. Bednarchik, “Electrical biasing of PZT ceramic material”, in *ISAF 2000. Proceedings of the 2000 12th IEEE International Symposium on Applications of Ferroelectrics (IEEE Cat. No. 00CH37076)*, IEEE, vol. 1, 2000, pp. 41–44.
- [200] L. Shaoping, W. Cao, R. Newnham, and L. Cross, “Electromechanical nonlinearity of ferroelectric ceramics and related non-180° domain wall motions”, *Ferroelectrics*, vol. 139, no. 1, pp. 25–49, 1993.
- [201] S. Priya, D. Viehland, A. V. Carazo, J. Ryu, and K. Uchino, “High-power resonant measurements of piezoelectric materials: Importance of elastic nonlinearities”, *Journal of applied physics*, vol. 90, no. 3, pp. 1469–1479, 2001.
- [202] S. Lee, D. Kang, Y. Je, and W. Moon, “Resonant frequency variations in a piezoelectric microcantilever sensor under varying operational conditions”, *Journal of Micromechanics and Microengineering*, vol. 22, no. 10, p. 105 035, 2012.
- [203] K. Uchino, H. Negishi, and T. Hirose, “Drive Voltage Dependence of Electromechanical Resonance in PLZT Piezoelectric Ceramics”, *Japanese Journal of Applied Physics*, vol. 28, no. S2, p. 47, 1989.

-
- [204] Y.-D. Kuang, G.-Q. Li, and C.-Y. Chen, “An admittance function of active piezoelectric elements bonded on a cracked beam”, *Journal of Sound and Vibration*, vol. 298, no. 1-2, pp. 393–403, 2006.
 - [205] F. Pigache and C. Nadal, “Modeling and Identification of Rosen-Type Transformer in Nonlinear Behavior”, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 12, pp. 2562–2570, 2011.
 - [206] S. Leadenham and A. Erturk, “Unified nonlinear electroelastic dynamics of a bi-morph piezoelectric cantilever for energy harvesting, sensing, and actuation”, *Nonlinear Dynamics*, vol. 79, pp. 1727–1743, 2015.
 - [207] J. W. Yi, W. Y. Shih, and W.-H. Shih, “Effect of length, width, and mode on the mass detection sensitivity of piezoelectric unimorph cantilevers”, *Journal of Applied Physics*, vol. 91, no. 3, pp. 1680–1686, 2002.
 - [208] M. Peddigari, M. S. Kwak, Y. Min, C.-W. Ahn, J.-J. Choi, B. D. Hahn, C. Choi, G.-T. Hwang, W.-H. Yoon, and J. Jang, “Lifetime estimation of single crystal macro-fiber composite-based piezoelectric energy harvesters using accelerated life testing”, *Nano Energy*, vol. 88, p. 106279, 2021.
 - [209] J. E. Sader, “Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope”, *Journal of Applied Physics*, vol. 84, no. 1, pp. 64–76, Jul. 1998.
 - [210] W. Y. Shih, X. Li, H. Gu, W.-H. Shih, and I. A. Aksay, “Simultaneous liquid viscosity and density determination with piezoelectric unimorph cantilevers”, *Journal of Applied Physics*, vol. 89, no. 2, pp. 1497–1505, Jan. 2001.
 - [211] T. Naik, E. K. Longmire, and S. C. Mantell, “Dynamic response of a cantilever in liquid near a solid wall”, *Sensors and Actuators A: Physical*, vol. 102, no. 3, pp. 240–254, 2003.
 - [212] R. Ohol and B. Vasuki, “Piezoelectric Actuated Stirrer for Solid Drug Powder-Liquid Mixing”, *Theoretical Foundations of Chemical Engineering*, vol. 56, no. 6, pp. 1100–1115, 2022.
 - [213] I. Solodov, N. Krohn, and G. Busse, “Nonlinear Ultrasonic NDT for Early Defect Recognition and Imaging”, in *European Conf. on NDT (ECNDT)*, Moscow, Citeseer Forest Grove, OR, USA, vol. 720, 2010.
 - [214] DIN EN 50160:2020-11, *Voltage characteristics of electricity supplied by public electricity networks; German version EN 50160:2010 + Cor.:2010 + A1:2015 + A2:2019 + A3:2019*. Deutsche Institut für Normung e.V. (DIN), Berlin, 2020.
 - [215] A. Arranz-Gimon, A. Zorita-Lamadrid, D. Morinigo-Sotelo, and O. Duque-Perez, “A review of total harmonic distortion factors for the measurement of harmonic and interharmonic pollution in modern power systems”, *Energies*, vol. 14, no. 20, p. 6467, 2021.
 - [216] B. D. Fulcher, M. A. Little, and N. S. Jones. “HCTSA Documentation”. Accessed: 03.2021. [Online]. Available: <https://hctsa-users.gitbook.io/hctsa-manual/>
 - [217] S. Petrovic, “A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters”, in *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, Citeseer, vol. 2006, 2006, pp. 53–64.

- [218] J. Xiao, J. Lu, and X. Li, “Davies Bouldin Index based hierarchical initialization K-means”, *Intelligent Data Analysis*, vol. 21, no. 6, pp. 1327–1338, 2017.
- [219] S. R. Paul and K. Y. Fung, “A generalized extreme studentized residual multiple-outlier-detection procedure in linear regression”, *Technometrics*, vol. 33, no. 3, pp. 339–348, 1991.
- [220] M. Ryu, G. Lee, and K. Lee, “Online sequential extreme studentized deviate tests for anomaly detection in streaming data with varying patterns”, *Cluster Computing*, vol. 24, no. 3, pp. 1975–1987, 2021.
- [221] B. J. Hamrock and W. J. Anderson, “Rolling-element bearings”, NASA Reference Publication, Tech. Rep., 1983.
- [222] F. Xu, N. Ding, N. Li, L. Liu, N. Hou, N. Xu, W. Guo, L. Tian, H. Xu, C.-M. L. Wu, et al., “A review of bearing failure Modes, mechanisms and causes”, *Engineering Failure Analysis*, p. 107518, 2023.
- [223] SKF PUB BU/13 14219 EN, “Bearing damage and failure analysis”, SKF Group, Tech. Rep., 2017.
- [224] IEEE Std 3006.8-2018, *IEEE Recommended Practice for Analyzing Reliability Data for Equipment Used in Industrial and Commercial Power Systems*. New York, NY: IEEE, 2018.
- [225] SKF, *SKF Explorer*, <https://www.skf.com/at/products/rolling-bearings/ball-bearings/deep-groove-ball-bearings/why-skf-explorer-deep-groove-ball-bearings>, Accessed: 08.05.2024.
- [226] ISO 281:2007-02(en), *Rolling bearings - Dynamic load ratings and rating life*. International Organization for Standardization (ISO), Geneva, Switzerland, 2007.
- [227] ISO 15243:2017(en), *Rolling bearings — Damage and failures — Terms, characteristics and causes*. International Organization for Standardization (ISO), Geneva, Switzerland, 2017.
- [228] SKF, *Bearing damage analysis: ISO 15243 is here to help you*, <https://evolution.skf.com/bearing-damage-analysis-iso-15243-is-here-to-help-you/>, Accessed: 08.05.2024.
- [229] J. Moubray, *Reliability-centered maintenance*. Industrial Press Inc., 2001.
- [230] P. Tchakoua, R. Wamkeue, T. A. Tameghe, and G. Ekemb, “A review of concepts and methods for wind turbines condition monitoring”, in *2013 World Congress on Computer and Information Technology (WCCIT)*, IEEE, 2013, pp. 1–9.
- [231] A. Hase, “Early detection and identification of fatigue damage in thrust ball bearings by an acoustic emission technique”, *Lubricants*, vol. 8, no. 3, p. 37, 2020.
- [232] ISO 13379-1:2012(en), *Condition monitoring and diagnostics of machines - Data interpretation and diagnostics techniques - Part 1: General guidelines*. International Organization for Standardization (ISO), Geneva, Switzerland, 2012.
- [233] ISO 13373-1:2002(en), *Condition monitoring and diagnostics of machines – Vibration condition monitoring - Part 1: General procedures*. International Organization for Standardization (ISO), Geneva, Switzerland, 2002.
- [234] R. B. Randall, *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. John Wiley & Sons, 2021.

-
- [235] ISO 13373-2:2016(en), *Condition monitoring and diagnostics of machines — Vibration condition monitoring - Part 2: Processing, analysis and presentation of vibration data*. International Organization for Standardization (ISO), Geneva, Switzerland, 2016.
 - [236] ISO 22096:2007(en), *Condition monitoring and diagnostics of machines — Acoustic emission*. International Organization for Standardization (ISO), Geneva, Switzerland, 2007.
 - [237] W. Zhou, T. G. Habetler, and R. G. Harley, “Bearing fault detection via stator current noise cancellation and statistical control”, *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4260–4269, 2008.
 - [238] X. Zhou, H. Zhang, X. Hao, X. Liao, and Q. Han, “Investigation on thermal behavior and temperature distribution of bearing inner and outer rings”, *Tribology International*, vol. 130, pp. 289–298, 2019.
 - [239] N. Bolander, H. Qiu, N. Eklund, E. Hindle, and T. Rosenfeld, “Physics-based remaining useful life prediction for aircraft engine bearing prognosis”, in *Annual Conference of the PHM Society*, vol. 1, 2009.
 - [240] G. J. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley Online Library, 2006, vol. 456.
 - [241] F. Ma, Z. Li, S. Qiu, B. Wu, and Q. An, “Transient thermal analysis of grease-lubricated spherical roller bearings”, *Tribology International*, vol. 93, pp. 115–123, 2016.
 - [242] T. A. Harris and M. N. Kotzalas, *Advanced Concepts of Bearing Technology: Rolling Bearing Analysis*. CRC press, 2006.
 - [243] J. Takabi and M. Khonsari, “Experimental testing and thermal analysis of ball bearings”, *Tribology International*, vol. 60, pp. 93–103, 2013.
 - [244] ISO 18434-1:2008(en), *Condition monitoring and diagnostics of machines — Thermography - Part 1: General procedures*. International Organization for Standardization (ISO), Geneva, Switzerland, 2008.
 - [245] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, “Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification”, in *PHM Society European Conference*, vol. 3, 2016.
 - [246] ISO 20958:2013(en), *Condition monitoring and diagnostics of machine systems — Thermography - Electrical signature analysis of three-phase induction motors*. International Organization for Standardization (ISO), Geneva, Switzerland, 2013.
 - [247] G. Kliman and J. Stein, “Methods of motor current signature analysis”, *Electric Machines and Power Systems*, vol. 20, no. 5, pp. 463–474, 1992.
 - [248] R. R. Schoen, T. G. Habetler, F. Kamran, and R. Bartfield, “Motor bearing damage detection using stator current monitoring”, *IEEE transactions on industry applications*, vol. 31, no. 6, pp. 1274–1279, 1995.
 - [249] M. J. Devaney and L. Eren, “Detecting motor bearing faults”, *IEEE Instrumentation & Measurement Magazine*, vol. 7, no. 4, pp. 30–50, 2004.

- [250] Society for Machinery Failure Prevention Technology. “Bearing Fault Data Set”. Accessed: 05.2024. [Online]. Available: <https://www.mfpt.org/fault-data-sets/>
- [251] Case Western Reserve University (CWRU). “Bearing Data Center”. Accessed: 05.2024. [Online]. Available: <https://engineering.case.edu/bearingdatacenter/download-data-file>
- [252] H. Huang and N. Baddour, “Bearing vibration data collected under time-varying rotational speed conditions”, *Data in Brief*, vol. 21, pp. 1745–1749, 2018, ISSN: 2352-3409. DOI: <https://doi.org/10.1016/j.dib.2018.11.019>
- [253] A. P. Daga, A. Fasana, S. Marchesiello, and L. Garibaldi, “The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data”, *Mechanical Systems and Signal Processing*, vol. 120, pp. 252–273, 2019.
- [254] A. Klausen, *UiA Accelerated Life Time Bearing Test Rig – Test 3, Variable speed around 50rpm*, Last viewed in: 05.2024, 2021. [Online]. Available: <https://doi.org/10.18710/NZJOIS>
- [255] S. Martin del Campo Barraza, F. Sandin, and D. Strömbergsson. “Dataset concerning the vibration signals from wind turbines in northern Sweden”. Accessed: 05.2024. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-70730>
- [256] VDI 2221 Part 1:2019 (en), *Design of technical products and systems - Model of product design*. Verein Deutscher Ingenieure e. V. (VDI), Düsseldorf, 2019.
- [257] G. Pahl, W. Beitz, J. Feldhusen, K.-H. Grote, et al., *Engineering Design: A Systematic Approach*. Springer, 1996, vol. 3.
- [258] B. Bender, K. Gericke, G. Pahl, W. Beitz, J. Feldhusen, and K.-H. Grote, *Pahl/Beitz Konstruktionslehre: Methoden und Anwendung erfolgreicher Produktentwicklung*, 9th ed., B. Bender and K. Gericke, Eds. Springer Berlin Heidelberg, 2021.
- [259] H. Kopka (Supervisor, O. K. Aimiyeagbon), *Erweiterung eines Prüfstands für die Zustandsüberwachung von Wälzlagern bei nichtstationären Betriebsbedingungen*. Paderborn University, Chair of Dynamics and Mechatronics (LDM), unpublished bachelor’s thesis, 2019.
- [260] H. Roloff and W. Matek, *Roloff-Matek Maschinenelemente: Tabellen*. Springer Science & Business Media, 2007, vol. 2.
- [261] N. Kumar et al., “Emerging Trends in Energy Storage Systems and Industrial Applications”, 2022.
- [262] T. Chen and C. Liaw, “Vibration acceleration control of an inverter-fed electrodynamic shaker”, *IEEE/ASME Transactions On Mechatronics*, vol. 4, no. 1, pp. 60–70, 1999.
- [263] G. F. Lang, D. Snyder, et al., “Understanding the Physics of Electrodynamic Shaker Performance”, *Sound and Vibration*, vol. 35, no. 10, pp. 24–33, 2001.
- [264] R. Fischer, H. Linse, R. Fischer, and H. Linse, “Elektrische Messtechnik”, *Elektrotechnik für Maschinenbauer: mit Elektronik, elektrischer Messtechnik, elektrischen Antrieben und Steuerungstechnik*, pp. 223–264, 2012.
- [265] G. S. Hegde, *Mechatronics*. Jones & Bartlett Learning, 2010.

-
- [266] A. Dietrich, “Electric rod actuators vs. hydraulic cylinders: A comparison of the pros and cons of each technology”, *Mach Des*, pp. 1–13, 2016.
 - [267] J. Pustavrh, M. Hočevár, P. Podržaj, A. Trajkovski, and F. Majdič, “Comparison of hydraulic, pneumatic and electric linear actuation systems”, *Scientific reports*, vol. 13, no. 1, p. 20 938, 2023.
 - [268] A. Tanenbaum and D. Wetherall, *Computer Networks*. Pearson Prentice Hall, 2011, ISBN: 9780132126953.
 - [269] National Instruments Corp., *National Instruments CompactDAQ cDAQ-9189*, <https://www.ni.com/en-us/shop/model/cdaq-9189.html>, Accessed: 31.05.2024.
 - [270] D. Ashlock and A. Warren, “The Engineer’s Guide to Signal Conditioning”, National Instruments Corp., White Paper, 2015.
 - [271] L. F. Villa, A. Reñones, J. R. Perán, and L. J. de Miguel, “Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load”, *Mechanical Systems and Signal Processing*, vol. 29, pp. 436–446, 2012.
 - [272] K. Medjaher, D. A. Tobon-Mejia, and N. Zerhouni, “Remaining useful life estimation of critical components with application to bearings”, *IEEE Transactions on Reliability*, vol. 61, no. 2, pp. 292–302, 2012.
 - [273] J. H. Kotecha and P. M. Djuric, “Gibbs sampling approach for generation of truncated multivariate gaussian random variables”, in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP99 (Cat. No. 99CH36258)*, IEEE, vol. 3, 1999, pp. 1757–1760.
 - [274] R. Krenek, J. Cha, and B. R. Cho, “Development of the convolutions of truncated normal random variables with three different quality characteristics in engineering applications”, *Computers & Industrial Engineering*, vol. 94, pp. 125–137, 2016.
 - [275] DIN ISO 10816-3:2018-01(de), *Mechanical vibration - Evaluation of machine vibration by measurements on non-rotating parts - Part 3: Industrial machines with nominal power above 15 kW and nominal speeds between 120 r/min and 15000 r/min when measured in situ (ISO 10816-3:2009 + Amd.1:2017)*. Deutsche Institut für Normung e. V. (DIN), Berlin, 2018.
 - [276] M. Behzad, S. Feizhoseini, H. A. Arghand, A. Davoodabadi, and D. Mba, “Failure threshold determination of rolling element bearings using vibration fluctuation and failure modes”, *Applied Sciences*, vol. 11, no. 1, p. 160, 2020.
 - [277] B. Wang, Y. Lei, N. Li, and N. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings”, *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401–412, 2018.
 - [278] T. Williams, X. Ribadeneira, S. Billington, and T. Kurfess, “Rolling element bearing diagnostics in run-to-failure lifetime testing”, *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 979–993, 2001.
 - [279] W. Wang, “A model to predict the residual life of rolling element bearings given monitored condition information to date”, *IMA Journal of Management Mathematics*, vol. 13, no. 1, pp. 3–16, 2002.

- [280] H. Qiu, J. Lee, J. Lin, and G. Yu, “Robust performance degradation assessment methods for enhanced rolling element bearing prognostics”, *Advanced Engineering Informatics*, vol. 17, no. 3-4, pp. 127–140, 2003.
- [281] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, “Residual life predictions from vibration-based degradation signals: A neural network approach”, *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694–700, 2004.
- [282] M. N. Kotzalas and T. A. Harris, “Fatigue Failure Progression in Ball Bearings”, *Journal of Tribology*, vol. 123, no. 2, pp. 238–242, 2001.
- [283] A. Bender, L. Schinke, and W. Sextro, “Remaining useful lifetime prediction based on adaptive failure thresholds”, in *Proceedings of the 29th European Safety and Reliability Conference (ESREL2019)*, 2019.
- [284] J. K. Choi et al., “Thermal characteristics of the spindle bearing system with a gear located on the bearing span”, *International Journal of Machine Tools and Manufacture*, vol. 38, no. 9, pp. 1017–1030, 1998.
- [285] Y. Dong, F. Chen, and M. Qiu, “Thermal-induced influences considered spindle unit angular contact ball bearing preload determination using embedded fiber Bragg gating sensors”, *International Journal of Distributed Sensor Networks*, vol. 18, no. 3, 2022.
- [286] J. Asama, T. Oi, T. Oiwa, and A. Chiba, “Simple driving method for a 2-DOF controlled bearingless motor using one three-phase inverter”, *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 4365–4376, 2018.
- [287] M. Kanemaru, M. Tsukima, T. Miyauchi, and K. Hayashi, “Bearing fault detection in induction machine based on stator current spectrum monitoring”, *IEEE Journal of Industry Applications*, vol. 7, no. 3, pp. 282–288, 2018.
- [288] E. V. Zaretsky, J. V. Poplawski, and L. E. Root, “Reexamination of ball-race conformity effects on ball bearing life”, *Tribology Transactions*, vol. 50, no. 3, pp. 336–349, 2007.
- [289] W. Wang, B. T. Kuhnell, and M. Harrap, “The Effects of System Characteristics on the Diagnosis of Bearing Defects”, *10th International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management (COMADAEM '97)*, vol. 1, p. 253, 1997.
- [290] O. K. Aimiyeagbon, *Run-to-failure data set of ball bearings subjected to time-varying load and speed conditions*, Data set, 2024. [Online]. Available: <https://doi.org/10.5281/zenodo.10805042>
- [291] R. Yang, Y. Jin, L. Hou, and Y. Chen, “Study for ball bearing outer race characteristic defect frequency based on nonlinear dynamics analysis”, *Nonlinear Dynamics*, vol. 90, pp. 781–796, 2017.
- [292] M. Xu, G. Feng, Q. He, F. Gu, and A. Ball, “Vibration characteristics of rolling element bearings with different radial clearances for condition monitoring of wind turbine”, *Applied Sciences*, vol. 10, no. 14, p. 4731, 2020.
- [293] M. Xu, Y. Han, X. Sun, Y. Shao, F. Gu, and A. D. Ball, “Vibration characteristics and condition monitoring of internal radial clearance within a ball bearing in a gear-shaft-bearing system”, *Mechanical Systems and Signal Processing*, vol. 165, p. 108 280, 2022.

- [294] Y. Reposo, “Use of pNNx statistics in the evaluation of heart rate variability at rest and during exercise”, *Archivos de Medicina del Deporte*, vol. 27, no. 138, pp. 255–259, 2010.

Appendix

A Characteristics and further results of the piezoelectric bending actuators

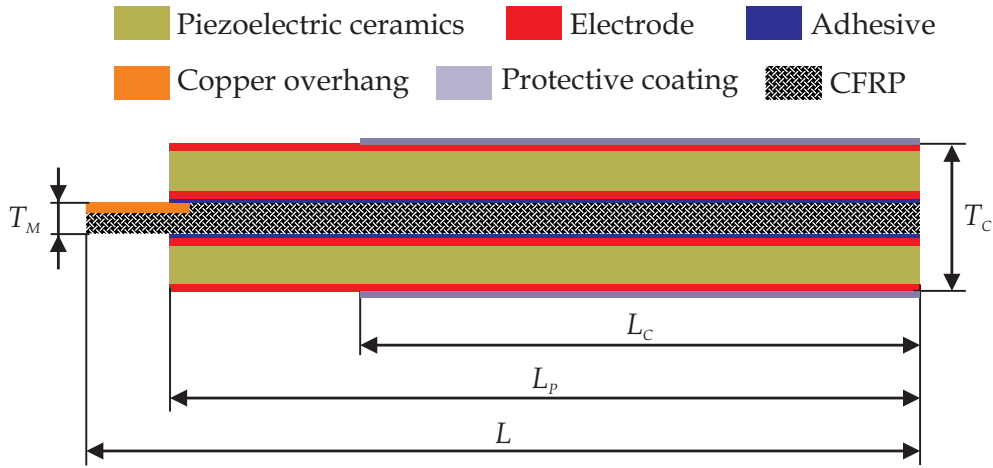


Figure .1: Side view showing the dimensions of the piezoelectric bending actuator adapted from [175], [178].

Table 1: Characteristics of the piezoelectric bending actuator from Johnson Matthey (Type1 - 427.0085.11F)

Attribute	Value
Length L	50.00 mm $\pm$ 0.3 mm
Length of coating L_C	38.20 mm $\pm$ 1.00 mm
Length of ceramic L_P	45.00 mm $\pm$ 0.50 mm
Width W	7.20 mm $\pm$ 0.05 mm
Thickness, coated T_C	0.81 mm $\pm$ 0.04 mm
Thickness CFRP (inclusive Copper overhang) T_M	0.28 mm $\pm$ 0.1 mm
Total deflection (at 230 V and 38 mm free length)	1.70 mm $\pm$ 0.40 mm
Capacitance per ceramic layer (at 1 kHz and 1 V)	45 nF $\pm$ 20%

A.1 Performance evaluation of the prognostics approaches

The resulting nRMSE values for the experiments P01 through P11 are presented in Table 2 for the three approaches described in Section 4.5. In Table 2, approach B represents

the ensemble RUL model utilizing solely the features derived from the current signal. Approach C considers voltage-scaled current features as input data to estimate the RUL, while approach E incorporates voltage features in addition to the current features.

As shown in Table 2, with the exception of experiment P06, approach C underperforms compared to the other approaches. Approach E demonstrates superior performance over both approaches B and C in more than 70% of the experiments. The evaluation results based on the nRMSE support the preceding observation based on the wMAPE that approach E, where the voltage features are utilized in the addition to the current features, achieves improved performance.

Table 2: Performance evaluation with nRMSE metric for the three described approaches

Experiment	nRMSE		
	Approach B	Approach C	Approach E
P01	10.29	10.40	7.30
P02	1.85	1.86	1.88
P03	1.24	2.33	2.35
P04	4.18	5.99	7.19
P05	7.74	9.62	3.22
P06	2.90	0.90	1.70
P07	37.33	38.82	22.65
P08	16.83	18.72	15.56
P09	2.41	2.16	1.61
P10	6.44	4.00	3.30
P11	2.83	5.66	2.52

B Characteristics and further results of the ball bearings

B.1 Dimensions of test ball bearing and support spherical roller bearing

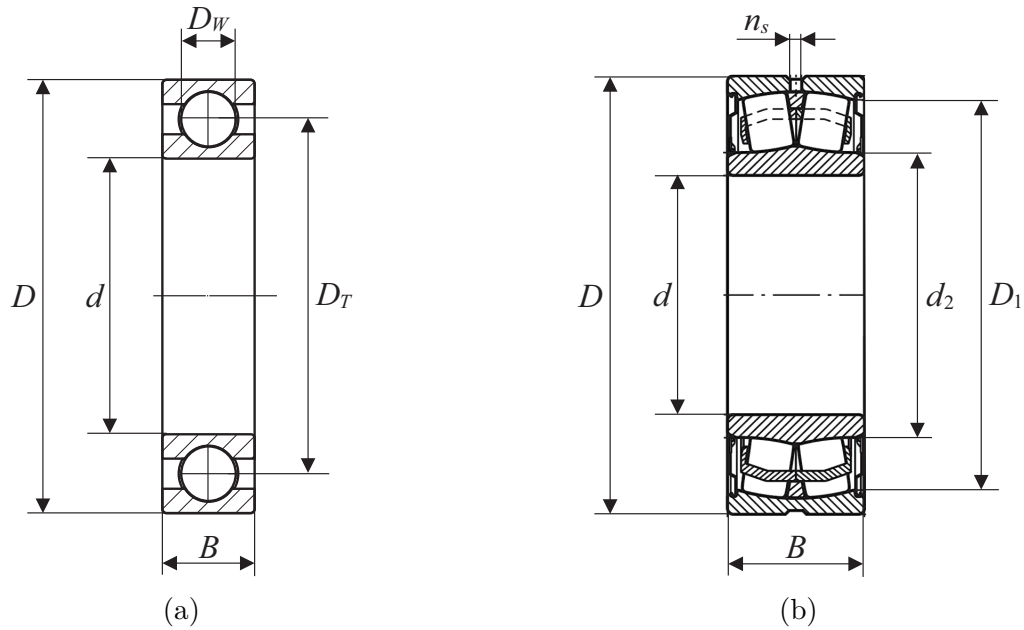


Figure .2: Bearing dimensions: (a) test ball bearing and (b) support spherical roller bearing

Table 3: Characteristics of the test bearing of type 61806-2RS

Attribute	Value
Bore diameter d	30 mm
Outer diameter D	42 mm
Width B	7 mm
Rolling circle diameter D_T	35.5 mm
Rolling element diameter D_w	3.18 mm
Number of balls z	19
Seal	NBR+08
Basic static load rating C_0	3.15 kN
Basic dynamic load rating C	4.00 kN
Limiting speed	12 000 rpm

Table 4: Characteristics of the support bearings of type WS22205-E1-XL-2RS

Attribute	Value
Bore diameter d	25 mm
Outer diameter D	52 mm
Width B	23 mm
Outer ring bore diameter D_1	46.1 mm
Inner ring raceway diameter d_2	29.8 mm
Basic static load rating C_0	42.5 kN
Basic dynamic load rating C	48.5 kN
Limiting speed	3600 rpm

B.2 Further performance evaluation results of the prognostics approaches

In Tables 6 and 7, approach L represents the bearing rating life model without considering the modification factor, while approach LT considers the bearing rating life model with temperature-dependent modification factor. Approaches B, BT, C and E utilize the exponential GPR model with ten condition-invariant features, while considering different levels of operating condition information, as described in Section 5.5. Approach B does not consider any further information, approach BT additionally considers the bearing relative temperature, approach C considers speed-scaled features, approach E considers the speed information in addition to the condition invariant features, while approach LT_E combines the benefits of approach LT and approach E via a 30th percentile of both approaches.

As presented in Tables 6 and 7 for the wMAPE and nRMSE, respectively, there is no one approach that outperforms the others in all the considered bearings, owing to several factors, such as the large lifetime variations and the non-negligible effects of the time-varying operating conditions. However, even without additional information, approach B produces satisfactory results. Furthermore, approach LT_E, that is, the ensemble of approach LT and approach E stands out, while considering both metrics. On a closer look at both tables, it is obvious that the metric score of Model approach LT_E is similar to that of approach LT, because the estimation with approach LT begins at the BoL onward and has a significant impact, as opposed to approach E, where the RUL is estimated after the transition time. Nonetheless, from the preceding results, it can be preliminarily concluded that although no single model outperforms the other models in all bearings, the incorporation of significant operating condition information, the bearing temperature and condition-invariant features engineered from the vibration signals provide some benefit towards detecting faults and estimating the RUL of ball bearings.

Table 5: Data Set Characteristics [290]

Experiment	Approx. Duration / h	Number of Measurements	Threshold Criterion	Defect
B01	2	377	-	IR, OR, B
B02	6	1116	$E_{T,vibration}$	IR
B03	3	614	$E_{T,temperature}$	IR
B04	6	1114	$E_{T,vibration}$	IR, B
B05	3	572	-	B
B06	67	12965	$E_{T,vibration}$	IR, B
B07	37	6393	$E_{T,temperature}$	IR
B08	11	1827	$E_{T,temperature}$	IR, B
B09	13	1813	$E_{T,vibration}$	OR, B
B10	19	3224	$E_{T,vibration}$	OR, B
B11	8	1953	$E_{T,vibration}$	IR, OR, B
B12	3	767	$E_{T,vibration}$	IR, B
B13	66	19417	$E_{T,vibration}$	IR, B
B14	41	12317	$E_{T,vibration}$	B
B15	76	22567	$E_{T,temperature}$	IR, B
B16	20	5891	-	IR, OR, B
B17	17	2733	$E_{T,vibration}$	B

Where $E_{T,vibration}$ and $E_{T,temperature}$ denote the vibration and threshold value, respectively. IR denotes inner race defect(s), OR denotes outer race defect(s) and B denotes rolling element defect(s).

Table 6: Performance evaluation with the wMAPE metric exemplarily for seven models

Bearing	Approach L	Approach LT	Approach B	Approach BT	Approach C	Approach E	Approach LT_E
B01	1.98	1.54	4.05	4.00	3.96	1.28	1.16
B02	0.37	0.28	2.78	3.98	3.92	3.77	0.28
B03	0.60	0.33	11.57	0.43	9.91	11.85	0.32
B04	0.52	0.87	2.71	2.87	4.93	2.84	0.74
B05	2.21	1.05	15.05	2.95	9.06	8.31	1.05
B06	1.89	1.61	0.69	0.80	0.80	0.82	1.61
B07	1.65	1.59	0.53	0.77	0.84	0.83	1.58
B08	0.91	1.40	1.76	1.12	1.20	0.94	1.32
B09	1.04	1.04	1.04	1.47	1.29	1.33	1.03
B10	1.85	1.82	0.80	0.92	0.94	0.88	1.79
B11	0.39	0.33	0.54	2.72	1.61	0.79	0.39
B12	3.95	5.00	1.98	2.58	3.10	3.21	2.80
B13	1.75	0.56	0.84	0.64	0.68	0.68	0.54
B14	1.58	0.57	1.04	0.74	0.94	1.39	0.55
B15	1.77	0.64	0.64	0.52	0.68	0.71	0.63
B16	1.23	1.08	11.05	3.96	2.53	2.05	1.07
B17	0.73	4.53	2.57	1.86	1.74	2.58	4.52

Table 7: Performance evaluation with the nRMSE metric exemplarily for seven models

Bearing	Approach L	Approach LT	Approach B	Approach BT	Approach C	Approach E	Approach LT_E
B01	4.72	3.77	18.31	14.93	13.80	4.99	3.41
B02	0.84	0.87	6.40	8.74	9.54	8.52	0.71
B03	1.41	0.89	25.71	1.10	21.98	26.52	0.93
B04	1.28	4.37	8.25	8.94	13.00	7.93	4.31
B05	5.22	3.02	34.71	11.97	22.31	21.21	3.05
B06	3.30	2.85	1.47	1.66	1.63	1.68	2.84
B07	2.92	2.81	1.24	1.83	1.76	1.77	2.76
B08	1.75	2.64	4.51	2.62	2.51	2.06	2.15
B09	1.97	1.97	3.33	5.49	3.49	4.63	1.62
B10	3.30	3.25	1.74	1.89	1.95	1.87	3.06
B11	0.82	0.75	1.59	8.00	4.24	1.80	0.72
B12	8.09	10.33	5.22	7.60	8.67	8.68	8.66
B13	3.04	1.23	1.82	1.40	1.46	1.46	0.96
B14	2.76	1.28	1.93	1.44	1.80	2.82	1.23
B15	3.08	1.35	1.39	1.10	1.46	1.51	1.16
B16	2.18	1.92	20.43	7.59	4.81	3.81	1.84
B17	1.43	9.97	5.23	4.08	3.70	5.24	9.96

Erklärung zur Zitation von Inhalten aus studentischen Arbeiten

In Ergänzung zu meinem Antrag auf Zulassung zur Promotion in der Fakultät für Maschinenbau der Universität Paderborn erkläre ich gemäß §11 der Promotionsordnung und unter Beachtung der Regelung zur Zitation studentischer Arbeiten:

Die von mir vorgelegte Dissertation habe ich selbstständig verfasst, und ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Es sind Inhalte studentischen Ursprungs (studentische Arbeiten) in dieser Dissertation enthalten.

Ich habe die verwendeten Arbeiten entsprechend der Regelung „Zitation aus studentischen Arbeiten in Dissertationen“ zitiert.

Paderborn, 10.07.2026

Osarenren Kennedy Aimiyeagbon, M.Sc.