Ecole Doctorale IAEM (Informatique, Automatique, Electronique-Electrotechnique, Mathématiques et Sciences de l'Architecture)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention : «Automatique, Traitement du signal et des images, Génie informatique»

Contributions à l'élaboration des modèles à partir de données pour l'estimation de la durée de vie résiduelle des roulements

Par

Fei Huang

le 27 Avril 2020

Membres du jury :

Rapporteurs :

Mme Zineb SIMEU ABAZI	Professeur, Université Grenoble Alpes, Grenoble
M. André BATAKO	Professeur, John Moores University, Liverpool

Examinateurs :

Mme Hind BRIL-EL HAOUZI	Professeur, Université de Lorraine, Epinal
M. Noureddine ZERHOUNI	Professeur, ENSMM, Besançon
M. Alexandre SAVA	Maître de Conférences HDR, Université de Lorraine, directeur de thèse
M. Kondo H. ADJALLAH	Professeur, Université de Lorraine, co-directeur de thèse

Laboratoire de Conception, Optimisation et Modélisation des Systèmes, LCOMS EA 7306
Université de Lorraine, Metz 57000, FRANCE

UNIVERSITÉ DE LORRAINE

Ecole Doctorale IAEM (Informatique, Automatique, Electronique-Electrotechnique, Mathématiques et Sciences de l'Architecture)

Doctoral Thesis

presented and defended publicly for obtaining the grade of

DOCTOR DEGREE OF THE UNIVERSITE DE LORRAINE

in « Automatics, Signal and Image Processing, Computer Engineering »

Contributions to data-driven model-based identification using small-size datasets for estimating bearings remaining useful life

by

Fei Huang

on 27 April 2020

Members of the jury :

Reviewers:

Mme Zineb SIMEU ABAZI	Professor, Université Grenoble Alpes, Grenoble
M. André BATAKO	Professor, John Moores University, Liverpool

Examiners:

Mme Hind BRIL-EL HAOUZI	Professor, Université de Lorraine, Epinal
M. Noureddine ZERHOUNI	Professor, ENSMM, Besançon
M. Alexandre SAVA	Associate Professor, Université de Lorraine, supervisor of the thesis
M. Kondo H. ADJALLAH	Professor, Université de Lorraine, co-supervisor of the thesis

Laboratoire de Conception, Optimisation et Modélisation des Systèmes, LCOMS EA 7306
Université de Lorraine, Metz 57000, FRANCE

Résumé étendu en français

En général, toute activité économique des entreprises manufacturières est basée sur des équipements qui sont soumis à un processus de dégradation dû à la charge qu'ils subissent lors de leur opération et/ou aux influences de l'environnement. Par conséquent, il est nécessaire de développer des stratégies à la fois pour surveiller leur capacité à accomplir les missions assignées, ainsi que pour maintenir ou rétablir la performance des équipements dégradés.

Les stratégies de maintenance élaborées pour assurer la disponibilité et le niveau de performance souhaité pour les équipements sont classifiées en trois catégories principales : maintenance corrective, maintenance préventive et maintenance prédictive.

Les actions de maintenance curative consistent à réparer l'équipement après l'arrivée d'une panne. Compte tenue caractère aléatoire des pannes, ces événements peuvent engendrer des arrêts de production importants et couteux. Ainsi, les stratégies de maintenance curative doivent être accompagnées des stocks importants de produits et de pièces de rechange afin de diminuer les temps de réparation et atténuer l'impact sur la production.

Les stratégies de maintenance préventive consistent à prévoir des actions de maintenance à de intervalles planifiées sans prendre en compte l'état de dégradation des composants à remplacer. Ces stratégies permettent une gestion plus simple des ressources pour la maintenance. Par contre, elles peuvent engendrer de la sur-maintenance, des arrêtes non nécessaires de production et une surconsommation de pièces de rechange.

Les stratégies de maintenance prédictive visent à programmer les opérations de maintenance selon l'état de santé des équipements, estimé sur la base des données collectées pendant leur surveillance. Une stratégie de maintenance prédictive comporte quatre étapes principales : l'acquisition de données, l'analyse de données et l'élaboration d'indicateurs et la prise de décision. L'étape d'acquisition de données vise à collecter et à stocker les données pertinentes par rapport à l'état de santé du système. L'information de décision est extraite à partir des données collectées durant l'étape d'analyse des données afin de comprendre l'évolution du processus de dégradation de l'équipement et d'estimer son état de santé. L'étape de prise de décision vise à choisir et à mettre en œuvre les actions de maintenance appropriées afin de garantir le niveau de performance souhaité pour l'équipement.

La maintenance prédictive repose sur la mise en œuvre de méthodes de pronostic, qui visent à prédire l'évolution de l'état de dégradation du système et à anticiper les pannes afin d'agir avant

leur arrivée. Vu sa capacité à anticiper les pannes et à éviter des arrêts de production couteux, la maintenance prédictive présente un intérêt économique considérable. Un indicateur important pour la mise en œuvre de cette stratégie de maintenance est la durée de vie résiduelle des équipements (RUL).

Les roulements sont des composants importants dans les équipements rotatifs et leur état de santé a un impact important sur la performance globale de ces systèmes. Une dégradation sévère des roulements peut engendrer des coûts importants pour un système de production. En effet, plus de 80% des roulements ont une durée de vie inférieure à la durée prescrite par les constructeurs, et 50% des pannes dans les machines tournantes sont engendrées par la dégradation de roulements.

Les travaux de recherche que nous avons réalisés dans cette thèse de doctorat portent sur l'estimation de la durée de vie résiduelle des roulements dans le cas où les données historiques sont disponibles seulement pour un petit nombre de roulements. Dans la pratique, on rencontre cette situation dans les cas des nouvelles machines, où peu de données liées à leur processus de dégradation sont disponibles. Nous avons apporté les contributions suivantes dans ce travail.

Dans un premier temps nous avons étudié les signatures temporelles communes extraites des signaux de vibration pour la caractérisation du processus de dégradation des roulements. Les signatures habituelles extraites des signaux de vibrations sont sensibles surtout dans la dernière étape du processus de dégradation. Suite aux limites constatés pour ces indicateurs, nous avons proposé de nouveaux indicateurs pour la surveillance de la dégradation des roulements. Ceux-ci ont la particularité d'être monotones et incluent l'information historique sur le processus de dégradation.

Par la suite, nous avons élaboré une méthode pour l'estimation de la durée de vie résiduelle des roulements sur la base des signatures extraites des signaux de vibration. Cette méthode est basée sur un système d'inférence flou proposé par Takaki et Sugeno (T-S FIS). Nous avons élaboré une technique d'identification des paramètres du modèle, à savoir les ensembles flous et les paramètres de sortie du modèle flou, à partir des signatures extraites des signaux de vibration. Dans nos travaux nous avons utilisé des signaux collectés pour un petit nombre de roulements identiques durant un cycle de vie complet.

Le processus de dégradation des roulements est complexe et leur durée de vie présente une dispersion importante même pour des roulements identiques, soumis à des charges identiques. L'utilisation d'un ensemble de jeux de données d'apprentissage de petite taille ne permettant pas d'estimer les paramètres du modèle avec une grande précision, les paramètres du modèle doivent être améliorés lorsque des données supplémentaires deviennent disponibles. Par

conséquent, nous avons élaboré une méthode pour la mise à jour des paramètres du modèle lorsque de nouvelles données sont rendues disponibles grâce à un processus d'optimisation itératif. Cette méthode utilise le paramétrage initial du modèle ainsi que les nouvelles données d'apprentissage. Nous avons élaboré un algorithme itératif basé sur la méthode du maximum de vraisemblance pour la mise à jour des paramètres du modèle T-S FIS.
Cette thèse est structurée en cinq chapitres.

Le premier chapitre résume l'étude bibliographique que nous avons effectué concernant les approches de surveillance de roulements et l'estimation de leur durée de vie résiduelle. Dans un premier temps, nous proposons une brève présentation des roulements et de leurs modes de défaillance. Nous présentons également les principales techniques d'acquisition de données et méthodes de traitement de signal pour la surveillance de l'état des roulements. Par la suite, nous proposons une étude comparative des principales techniques présentes dans la littérature pour l'estimation de la durée de vie résiduelle des roulements. Ces approches peuvent être classées en trois catégories : 1) approches basées sur un modèle analytique du processus de dégradation ; 2) approches basées sur une modélisation statistique et 3) approches basées sur l'intelligence artificielle.
Les approches basées des modèles analytique considèrent que le processus de dégradation est bien connu et qu'il peut être décrit avec précision par un modèle physique. Ces approches ne nécessitent pas un volume important de données de surveillance. Cependant, le processus de dégradation des roulements est très complexe et il ne peut pas être décrit facilement par un modèle analytique.
Les approches basées sur une modélisation statistique s'affranchissent de modèle analytique, mais elles nécessitent une quantité importante de données historiques afin d'identifier les paramètres du modèle statistique.
Les approches basées sur des techniques d'intelligence artificielle ont la capacité d'intégrer des relations non-linéaires complexes entre les paramètres qui gouvernent le processus étudié. Ces approches sont utilisées lorsqu'il est difficile de construire un modèle analytique ou statistique approprié dont l'ordre est difficile à déterminer. Cependant, la performance des algorithmes d'intelligence artificielle est fortement influencée par la valeur de nombreux paramètres qui doivent êtres initialisés à travers un processus d'apprentissage qui nécessite un volume important de données d'entrainement.

Notre projet de recherche porte sur l'élaboration d'une méthodologie pour l'estimation de la durée de vie résiduelle des roulements lorsqu'un volume faible de données historiques est disponible. Par conséquent, nous avons focalisé notre attention sur la logique floue. Ce choix a été motivée par l'aptitude de cet outil à aborder des problèmes de décision lorsque l'information disponible est incomplète, voire imprécise.
La dernière section de ce chapitre présente les objectifs de recherche abordés dans le cadre de cette thèse de doctorat.

Le chapitre 2 porte sur l'extraction des signatures pour la surveillance du processus de dégradation et l'estimation de la durée de vie résiduelle de roulements. Nous avons orienté notre attention sur les signatures dans le domaine temporel d'abord, compte tenu de notre objectif initial qui est l'estimation de la durée de vie résiduelle. Dans un premier temps, nous avons étudié l'efficacité des signatures utilisées de manière courante pour la surveillance des roulements et nous avons montré leurs limitations. Par la suite nous avons élaboré deux algorithmes pour l'extraction de nouvelles signatures à partir des signaux de vibration. Ces algorithmes sont basés sur une méthode de représentation linéaire par morceaux, la segmentation des séries temporelles et l'analyse par composantes principales (PCA) pour élaborer des signatures qui évoluent de manière quasi-monotone avec le processus de dégradation des roulements.
Les données utilisées, présentes dans la littérature, ont étés générées par une plateforme expérimentale par l'observation périodique des signaux de vibrations sur un cycle complet de fonctionnement pour des roulements identiques soumis à des charges identiques. Nous avons construit un vecteur de signatures basé sur un ensemble de signatures temporelles. Par la suite avons proposé deux fonctions coût qui quantifient l'écart d'information entre l'espace de données original et l'espace réduit, déterminé par la méthode ACP. Ces fonctions coût sont employées pour la segmentation de la série temporelle multivariée en un ensemble de segments caractérisés par une évolution linéaire. Chaque segment correspond à un état de dégradation.

Par la suite, dans le chapitre 3 nous présentons la méthode que nous avons élaboré pour estimer la durée de vie résiduelle des roulements à partir d'un jeu de données d'apprentissage de petite taille, collecté pour un petit nombre de roulements à travers des observations périodiques sur un cycle complet de fonctionnement. Nous avons élaboré une technique d'identification des paramètres du modèle T-S FIS basée sur l'analyse des distributions mixtes. Dans un premier temps, nous avons estimé les paramétrés des ensembles flous par une méthode d'agrégation

floue. Chaque ensemble flou obtenu correspond à un état de dégradation dans l'espace des signatures. Par la suite, nous avons utilisé la méthode du maximum de vraisemblance pour calculer les paramètres des clusters temporels et pour estimer une probabilité associée à chaque état de dégradation. Une méthode d'estimation basée sur les moindres carrées est utilisée pour identifier les paramètres de sortie du modèle.

Le chapitre 4 présente la méthode que nous avons élaboré pour la mise à jour des paramètres du modèle lorsque des données d'apprentissage supplémentaires deviennent disponibles suite à un processus d'accumulation de connaissances.
En effet, le processus de dégradation des roulements est complexe et leur durée de vie présente une dispersion importante même pour des roulements identiques, soumis à des charges identiques. Et l'utilisation d'un ensemble de jeux de données d'apprentissage de petite taille ne permet pas d'estimer les paramètres du modèle T-S FIS avec une grande précision. Par conséquent, les paramètres du modèle doivent être améliorés lorsque de nouvelles données supplémentaires deviennent disponibles.
Nous avons élaboré une méthode pour la mise à jour des paramètres du modèle lorsque des nouvelles données sont rendues disponibles suite à un processus de collecte de données adaptatif et d'amélioration itérative des paramètres du modèle. Cette méthode utilise les paramètres initiaux du modèle et les nouvelles données pour l'apprentissage. Nous proposons un algorithme itératif basé sur la méthode du maximum de vraisemblance pour la mise à jour des paramètres d'entrée du modèle T-S FIS. Nous faisons l'hypothèse que le nombre de règles reste inchangé. Par la suite, la méthode du maximum de vraisemblance est utilisée pour calculer les paramètres de sortie du modèle. L'équation linéaire obtenue est résolue de manière itérative par la méthode de Kaczmarz.
Des données fournies par un banc de test dans la littérature ont été utilisées pour évaluer les méthodes proposées. Les résultats numériques ont montré la tendance d'évolution monotone croissante des indicateurs proposés avec l'évolution du niveau de dégradation. De même, les méthodes élaborées pour l'identification et la mise à jour des paramètres ont été comparées à des méthodes existantes dans la littérature, avec des résultats prometteurs.
Les travaux de recherche réalisées dans le cadre de cette thèse de doctorat ont permis les contributions suivantes :

1) Élaboration de deux fonctions coût pour la segmentation des séries temporelles multivariées. Ces fonctions coût exploitent l'écart des informations dans l'espace original des signatures et l'espace réduit construit par la méthode ACP.

2) Construction de deux indicateurs pour suivre la dégradation des roulements à partir des signaux de vibration. Ces indicateurs évoluent de manière monotone avec la dégradation et mémorisent l'historique de ce processus.
3) Élaboration d'un modèle basé sur un système d'inférence flou du modèle de Tagaki-Sugeno (T-S FIS) pour l'estimation de la durée de vie résiduelle des roulements dans le cas particulier où les données d'apprentissage ont étés collectées sur un petit nombre de roulements identiques, soumis à des chargements identiques.
4) Élaboration d'une méthode d'amélioration et de mise à jour des paramètres du modèle T-S FIS, lorsqu'un de nouveaux jeux de données supplémentaires d'apprentissage devient disponible permettant ainsi une amélioration du modèle de dégardation.

Une perspective immédiate de ces travaux de recherche porte sur l'amélioration des méthodes proposées. Dans ce sens, nous envisageons d'étudier l'extension des indicateurs proposées à d'autres composants et l'utilisation d'indicateurs normalisés. Nous envisageons également de relaxer l'hypothèse selon laquelle le nombre de règles reste fixe pour la mise à jour des paramètres du modèle T-S FIS.

A plus long terme, nous envisageons d'orienter nos travaux sur le développement des approches de collecte adaptatifs de données pour la surveillance et le diagnostic dans une architectures de système à paramètres distribués.

Acknowledgments

I would like firstly to express my sincere gratitude to my main supervisor, Professor Alexandre SAVA, for his guidance and support throughout this research. It is his inspiration that encourages me to cope with all the faced challenges and gain rewarding research knowledge.

My sincere appreciation is extended to my co-supervisor, Professor Kondo Hloindo ADJALLAH. His inestimable guidance and warm-hearted help both in research and life support and encourage me a lot.

I appreciate the help of all the friends I have made in Metz. They give me a lot of help and advice in my life and research.

I would like to extend my special thanks to my parents, my wife, and my daughter for their understanding, support, and encouragement.

Finally, this research was financially funded by the scholarship from the Huaiyin Institute of Technology, China. I am sincerely grateful to the Huaiyin Institute of Technology for its support during the research.

Metz, France
January 1, 2020

Content

Figures

Tables

Abbreviations

AE	Approximate entropy
ANFIS	Adaptive Neuro-Fuzzy Inference Systems
ARRMSE	Average of RRMSE
ASDS	Average Standard Deviation of SRCC
BPGD-LSE	The least square estimation and backpropagation gradient descent
CBM	Condition Based Maintenance
CD	Correlation dimension
CPV	Cumulative Percent Variance
FCM	Fuzzy C-Means Algorithm
FIS	Fuzzy Inference System
FSC-LSE	The subtractive clustering and least square estimation
GA	Genetic algorithm
LLE	Largest Lyapunov exponent
MISO	multiple-input single-output
MLWLSE	The subtractive clustering with maximum likelihood and weighted least square estimation
PCA	Principal Component Analysis
PMR	Peak-Magnitude-to-RMS
PUL	Past Useful Life
RMS	Root Mean Square
RRMSE	Relative Root Mean Square Error
RUL	Remaining Useful Life
SA	Simulated annealing
SDHT2	Segmented Discarded Hotelling T-square
SE	Spectral entropy
SPE	Segmented Projected Error
SRCC	Spearman Rank Correlation Coefficient
STD	Standard Deviation
T-S	Takagi-Sugeno
$VSDHT^2$	Vector of Segmented Discarded Hotelling T-square
VSPE	Vector of Segmented Projected Error

Symbols

$k=1, 2, \ldots, K$	The label of sample intervals
F	The vector consists of multiple common features
$i=1, 2\ldots, I$	The variable label
$m=1,2,\ldots, M$	The segment label
TF	The full projected space Hotelling T square
TR	The reduced projected space Hotelling T square
Λ^2	The segmented discarded Hotelling T square
$s=1, 2, \ldots, S$	The label of the training bearings
$q=1, 2, \ldots, Q$	The label of variables in the indicator $VSDHT^2$
p	The number of retained principal components
Δ^2	The segmented projected error
$b=1, 2, \ldots, B$	the label of variables in the indicator VSPE
ρ	The consumed useful life ratio
V	The input vector of the model
r	The weight of rules
w	The degree of fulfillment of rules
μ	The input membership function of T-S FIS
R	Rules
θ	The regime in a system
$j=1, 2, \ldots, J$	The label of regimes/rules/component distributions
ς	The priori probability of component distributions/ degradation states
σ	The standard deviation associated with component distributions/ degradation states/ input membership function
c	The centroid associated with component distributions/ degradation states/ input membership function
a, b	The parameters of output membership function
Y	The clusters validity index
σ'	The tuned standard deviation associated with input membership function
c'	The tuned centroid associated with input membership function
a', b'	The tuned parameters of output membership function

General introduction

The available useful life has been an essential aspect of evaluating industrial equipment. With the development of technology, new techniques for increasing the reliability of industrial equipment have been developed. However, the products still deteriorate over time as they operate under certain pressure or load. Hence, monitoring is necessary to assess the ability of the equipment to accomplish their mission and maintenance is necessary to recover the performance of degraded equipment.

There are three main kinds of maintenance strategies for ensuring the adequate reliability of facilities: breakdown maintenance, planned or scheduled maintenance and condition-based maintenance (CBM) [1]. The breakdown maintenance, maintenance is employed when equipment needs to be repaired after a breakdown occurred. Its advantages are low-cost policy and minimal management. Obviously, its disadvantages are consequences of the fact that the equipment could be broken down at the most inconvenient times when there is a larger inventory of work in progress, causing high downtime and potentially important production lost. In order to reduce the repair time, both the spare parts and the maintenance crew need to be available. Therefore, a large inventory of spare parts is needed and a maintenance crew need to be oversized. This strategy is not efficient for an integrated plant. The planned maintenance strategy carried out maintenance actions only when the machine operating time achieved a predefined time. This strategy simplifies the management of maintenance resources. Its disadvantages result from the fixed predefined time, i.e., maintenance is carried out upon the predefined time rather than the condition of equipment while equipment degradations are diverse. This may lead to over maintenance, unnecessary downtime and oversupply of replacement elements.

Conditioned based maintenance is a maintenance program that manages maintenance based on the information collected through condition monitoring. Thus, the disadvantages of planned schedule maintenance can be avoided. It is state-of-the-art in maintenance strategies.

There are three key steps in a CBM program, i.e., data acquisition, data processing and maintenance decision-making [2]. The data acquisition step aims to collect and store the data relevant to the system health. Information is extracted from data during the data processing step in order to understand and analyze the degradation process and the condition of the equipment.

The maintenance decision-making step aims to recommend efficient maintenance policies to achieve the performance objectives assigned to the equipment.

There are two main categories of techniques for maintenance decision support in a CBM program: diagnosis and prognosis [3]. On the one hand, the diagnosis refers to detection, isolation and identification of faults when they occur. On the other hand, the prognosis attempts to predict faults or failures before they occur. The prognosis provides higher economic interest than diagnosis due to its ability to prevent faults or failures. Additionally, the prognosis allows remaining useful life (RUL) estimation.

Bearings are critical common units in rotary machinery. Their failure has a severe influence on the overall system performance. Unexpected bearing failure can result in costly breakdowns in manufacturing. Less than 20% of bearings reach their end of life, and around 50% of all rotating machinery is caused by bearing failures, which highlights the need to monitor the bearings and provide effective indicators and models to assist in the preventive maintenance decisions for economic and security reasons [4] [5]. Therefore, bearings have received much attention in the field of CBM.

In this research, we address the problem of bearing RUL estimation when historical data are available only for a small number of bearings. In practice, this situation appears with the introduction of a new type of bearings.

First, considering the limitation of vibration signal based standard features for monitoring the bearings degradation, which do not increase continuously with the degradation cumulation, the research aims to find effective ways to build new bearings degradation indicators to provide an efficient description of the bearings' condition. Since bearings are mechanical components and self-repair is improbable during the non-utilization periods. Namely, the degradation evolution of bearings is cumulative and increasing. It is natural that a feature or indicator with an underlying monotonic trend, which increases continuously with the degradation cumulation, can describe the degradation progression of bearings well.

Second, this research aims to elaborate a method to identify a classic simple structure model with a small number of parameters for estimating bearings RUL using bearings degradation indicators extracted from the vibration signals observed over a complete run-to-failure cycle from a small number of identical bearings operating under identical load.

Third, this research aims to find an effective method to tune the parameters of the model based on additional data sets from a small number of identical bearings operating under identical load.

To illustrate the contributions of this research work, the thesis is structured into five chapters.

Chapter 1, introduces a literature review on bearings, their failure modes and related prognosis methods. We provide a brief presentation of the data acquisition techniques, a survey of signal process techniques and a comparative study of the main existing bearings remaining useful life (RUL) estimation methods. Finally, the conclusion of this chapter and the research objects are provided.

Chapter 2 deals with the feature extraction for bearing degradation monitoring. The first subsection provides a brief presentation of the standard degradation features in this field. We highlight the effectiveness of standard-bearing condition monitoring features and dissuss their limitations. Piecewise linear representation method is introduced in the second subsection. The segmentation algorithm of time series is presented in the third subsection. Principal component analysis for multivariate analyses is introduced in the fourth subsection. In the fifth and sixth sections, we propose new degradation indicators for bearing condition monitoring. In the last subsection, the conclusions of this chapter as well as new degradation indicators are presented.

Furthermore, starting with Chapter 3, we focus on RUL estimation for bearings. In the first section, the modeling scheme is introduced. The second section provides a brief description of Takagi-Sugeno (T-S) fuzzy inference system (FIS). In the third section, a brief description of the fuzzy subtractive clustering method is provided. The maximum likelihood estimation of mixture distribution analysis is introduced in the fourth section. In the fifth section, the mixture distribution analysis based fuzzy model identification method is proposed. Concluding remarks are given in the last section of this chapter.

Chapter 4, in the first subsection, the problem statement is introduced. In the second subsection, the approach for tuning parameters of the T-S FIS model is proposed to offset the drawback of small-size samples for model tuning. This fuzzy model tuning approach relies on the mixture distribution analysis. Finally, the summary section concludes this chapter.

Then, Chapter 5, explores the performance of the tools worked out in this research through numerical simulation using a benchmark dataset.

The last chapter provides a general conclusion with some discussions and suggestions of further directions about this research work.

Chapter 1 - Bearing condition monitoring for remaining useful life assessment

This chapter provides the literature review on bearing condition monitoring aiming to assess the remaining useful life and introduces our research.

Bearings are critical standard components in rotary machinery and their degradation has a severe influence on the overall system performance. Unexpected bearing failure can result in costly breakdowns in manufacturing. Less than 20% of bearings reach their end of life, and around 50% of all rotating machinery is caused by bearing failures, which highlights the need to monitor the bearings and provide effective indicators and models to assist in the preventive maintenance decisions for economic and security reasons [4] [5]. Therefore, bearings have received much attention in the field of condition monitoring.

In general, the procedure of bearings condition monitoring consists of the following steps: data acquisition, data processing, and decision-making. The data acquisition is based both on sensors, which could transfer the information into electrical or other physical signals and on observations from human operators. As the second step, data processing aims to identify the physical and statistical parameters (known as indicators or features) that can reflect the degradation status of the bearings. Based on the results of data processing, the decision-making step aims to estimate the remaining useful life of the bearings, to implement an appropriate maintenance policy and/or to detect the failure and identify the root causes. Our research work concerns with bearing remaining useful life assessment. Therefore, we conducted our literature review mainly on this topic.

1.1 Background on bearing degradation

1.1.1 Bearings

The basic function of a bearing is to support, guide, and reduce the friction of an axle. Nowadays, industrial applications use several categories of bearings. The most important ones are plain bearings and rolling bearings. Plain bearings are the simplest type of bearings. They do not contain rolling elements and the axle moves in the opposite direction of the bearing on a sliding surface. In the rolling bearings, the motion and the forces transfer is performed by rolling elements which transfer motions and forces, resulting in reduced friction comparing with plain bearings [6].

There are two categories of rolling bearings: ball bearings and roller bearings. The ball bearings are the most common where the rolling elements are balls. Figure 1.1 shows a typical example of a ball bearing.

Ball bearings generally consist of four main basic parts: rolling elements, outer ring, inner ring, and cage [7]. Figure 1.2 shows the structure of the ball bearing. The outer ring is installed in the housing of the machine, which is stationary. The inner ring is installed on the axle of the machine, which rotates synchronously with the axle. The rolling elements separate the inner and outer ring and transmit the load of the bearing through a small surface with a thin lubricating film. The cage is responsible for separating adjacent rolling elements and for holding the bearing structure together.

Figure 1.1 A ball bearing [7].

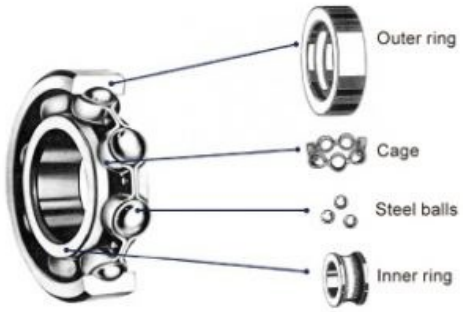


Figure 1.2 The structure of a ball bearing [8].

1.1.2 **Symptoms of bearings failures**

Bearings usually transmit the motions and the forces applied to the machine during operation [9]. Therefore, bearings' condition is critical for the performance of a production system.

There are several main origins of bearing failure: excessive loading, overheating, false brinelling, true brinelling, normal fatigue failure, reverse loading, contamination, lubricant failure, corrosion, misalignment, loose fits, tight fits [10]. Each of the origins can result in the formation of particular damage that corresponds to a specific mechanism and can leave its special imprint on the bearing. In the sequel, we provide a brief presentation of the main bearing degradation symptoms. A presentation of the symptoms for ball bearings failures is discussed in [10].

Severe spalling

Excessive loads can cause premature fatigue and generate ball wear paths, illustrated in Figure 1.3. Heavy ball wear paths, overheating and a widespread spalling result in a short bearing useful life [10].

Figure 1.3 Severe spalling and its location of ball bearing [10].

Discoloration

Overheating can cause a discoloration mark of the rings, balls, and cages from gold to blue, as illustrated in Figure 1.4. Lubricant failure could result in overheating [10]. The overheating results in high temperatures of bearings, exceeding 400°F, which can anneal the ring and ball materials. The resulting loss in hardness causes early failure. Balls and rings can deform in an extreme situation.

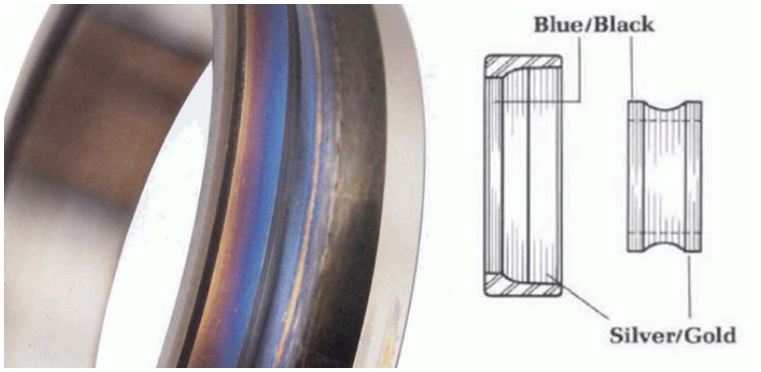


Figure 1.4 Discoloration and the location of ball bearing [10].

False Brinell marks

The false Brinell marks, illustrated in Figure 1.5, are elliptical wear marks in an axial direction at each ball position with a bright finish and sharp demarcation. A ring of brown debris often surrounds them and indicates excessive external vibration. A small relative motion between balls and raceway occurs in non-rotating ball bearings that are subject to external vibration. When the bearing is not turning, there is no oil formed. Hence the small relative motion between balls and raceway causes raceway wear. Wear debris will oxidize and accelerate the wear process [10].

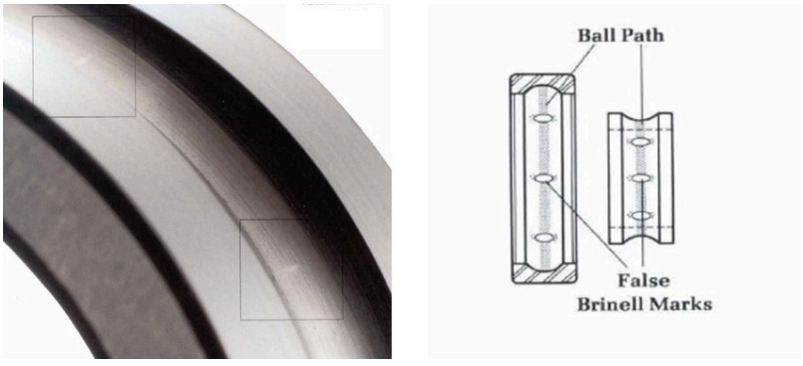


Figure 1.5 False Brinell marks and its location of ball bearing [10].

Brinell marks

Brinelling can appear because of any static overload or severe impact, e.g., using hammers to remove or install bearings, dropping or striking assembled equipment, and pressing a bearing

onto a shaft by applying force to the outer ring. Brinell marks, illustrated in Figure 1.6, are indentations in the raceways and increase bearing vibration (noise). Premature fatigue failure could be caused by severe Brinell marks [10].

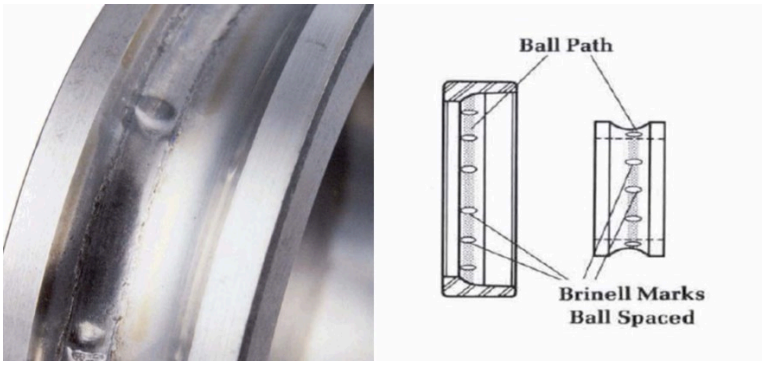


Figure 1.6 Brinell marks and the location of ball bearing [10].

Wear bands on balls

When the load is in the incorrect direction, the corresponding low shoulder of the outer ring truncates the elliptical contact area. This abnormal behavior generates excessive stress at a high temperature. Therefore, increased vibration and early failure will occur. The symptom is the wear bands on balls, illustrated in Figure 1.7. It shows a grooved wear band caused by the ball riding over the outer edge of the raceway [10].

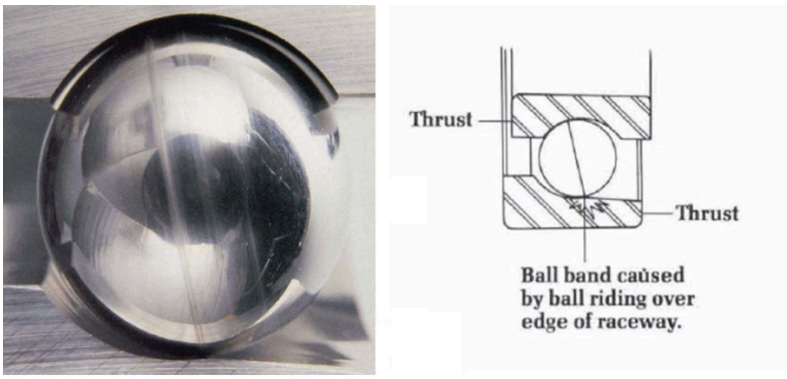


Figure 1.7 Wear band on balls and its location of ball bearing [10].

Denting of the bearing raceways and balls

Denting of the bearing raceways and balls, illustrated in Figure 1.8, appears due to contamination and results in high vibration and wear. Contaminants could be airborne dust, dirt or any abrasive substance that finds its way into the bearing. They can come from dirty tools, contaminated work areas, dirty hands, and foreign matter in lubricants or cleaning solutions [10].

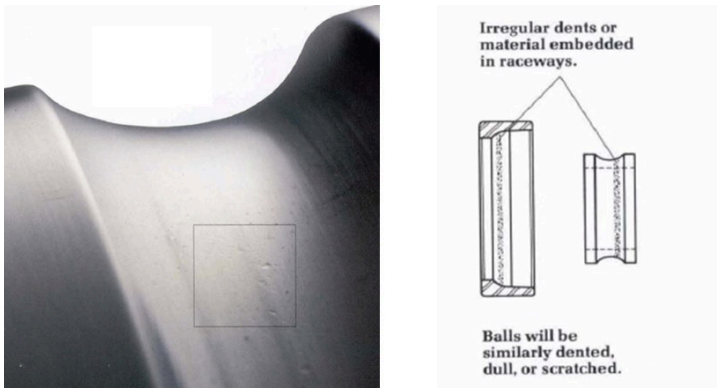


Figure 1.8 Denting and the location of ball bearing [10].

Stains and deposits

Stains and deposits, illustrated in Figure 1.9, are a consequence of corrosion due to exposing bearings to corrosive fluids or to a corrosive atmosphere. They are represented by red or brown areas on balls, raceways, cages, or bands of ball bearings. The consequences are the increase in vibration and radial clearance or loss of preload. The corrosion can initiate early fatigue failures in extreme cases [10].

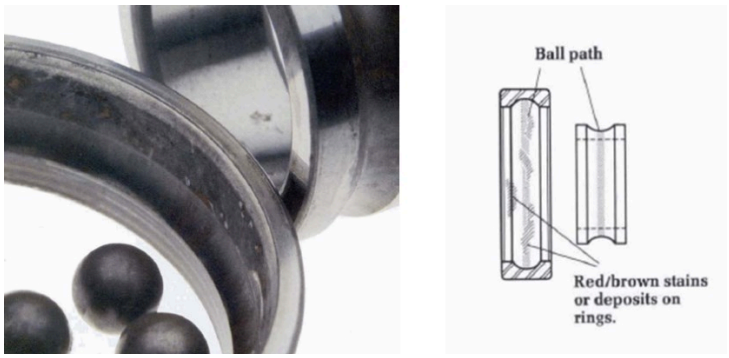


Figure 1.9 Stains and deposits and the location of ball bearing [10].

Misaligned wear path

The misaligned wear path, illustrated in Figure 1.10, is a ball wear path that is not parallel to the raceway edges. In general, the misalignment can be caused by bent shafts, burrs or dirt on shaft or housing shoulders, shaft threads that are not square with shaft's seats, and locking nuts with faces that are nor square to the thread axis. There could be an abnormal temperature rise in the bearing and housing. Meanwhile, heavy wear could occur in cage ball pockets [10].

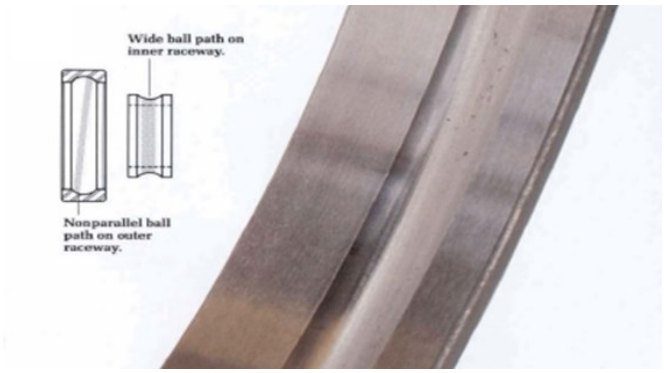


Figure 1.10 Misaligned wear path and its location of ball bearing [10].

Fretting color

Loose fits result in relative motion between mating parts. When the relative motion between mating parts is slight but continuous, fretting occurs, as illustrated in Figure 1.11. The fine metal particles which oxidize arises through the fretting, which show as a distinctive brown color called fretting color [10].

These fine metal particles are abrasive and aggravate the looseness. When the looseness increases enough to cause considerable movement of the inner or outer ring, the mounting surfaces (bores, outer diameters, face) wear increases and generates heat, noise, and runout problems [10].

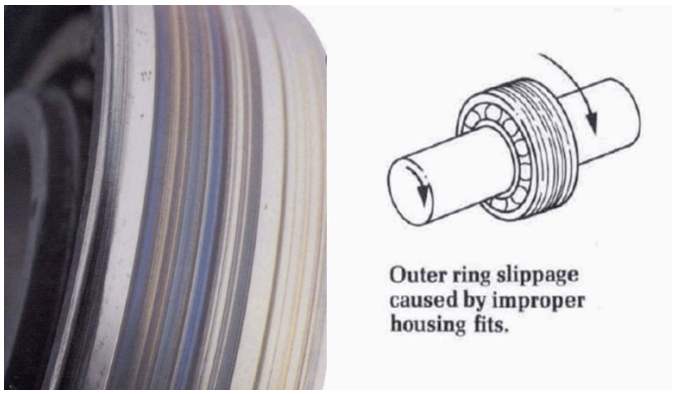


Figure 1.11 Fretting color and its location of ball bearing [10].

Tight fit wear path

Tight fit could cause heavy ball wear paths in the bottom of the raceway around the entire circumference of the inner ring and the outer ring, as illustrated in Figure 1.12. The balls will become excessively loaded when interference fits exceed the radial clearance at operating

temperature. A rapid temperature rise accompanied by high torque occurs in this case, generating rapid wear and fatigue [10].

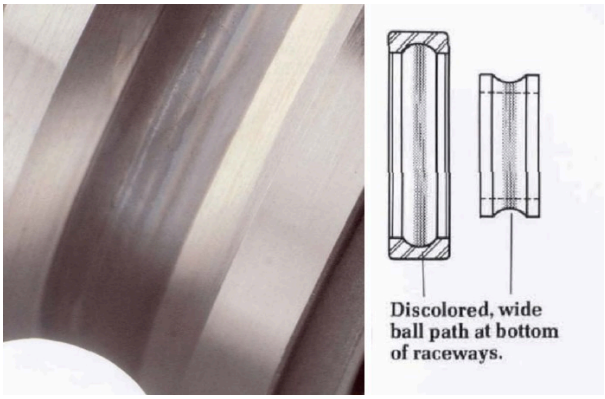


Figure 1.12 Tight fit wear path and its location of ball bearing [10].

1.2 Data acquisition for bearing condition monitoring

The data acquisition process aims to collect and store useful data (information) from targeted physical devices. This step is essential in systems condition monitoring, faults diagnosis, and prognostics.

In general, there are two categories of collected data [3]: event data and condition monitoring data. The event data gathers the information of targeted physical device considering the past events, e.g., installation, breakdown, overhaul, minor repair, preventive maintenance, oil change, etc. The condition monitoring data concerns the measurements concerning the health condition or the state of the physical device.

The increasing computing power and advanced sensor technologies enabled more powerful and less expensive data acquisition technologies. Condition monitoring data acquisition for bearings becomes more affordable and feasible. There are several types of sensors elaborated in [11].

Proximity probes based on eddy current, strain gauge, and fiber optic sensors are used to measure outer ring deflection at each ball pass by being installed on the outer ring [12]. However, to obtain a clear view of the bearing's outer ring, sensors for proximity probes need refitting for installation into the bearing housing. This data collection technique is challenging to implement.

Velocity sensors can measure the low frequency response of machines. However, their ability to work is only in a narrow frequency range affected their popularity [13].

Lasers sensors and microphones are used for contactless measurement [14]. They use the principle of Doppler shifting to measure the surface velocity of machines. During the measurement process, obstacles must not exist between the bearings and sensors. Microphones measure the acoustical response of machines effectively only in specific cases [14].

Acoustic emission sensors detect transient elastic waves generated from a rapid release of strain energy caused by deformation or damage within or on the surface of a material [15]. Acoustic emission sensors are useful for nondestructive testing measurement. However, their performance suffers in the presence of audible noises. In practice, the difficulty for calibrating, for performing repeatable and transferable measurements, and the cost of acoustic emission hardware, are essential obstacles to use acoustic emission sensors [16].

Accelerometers can provide a wide dynamic range and a wide frequency range for vibration measurement. They allow collecting bearing vibration signals easily. However, there is a need for a relatively rigid connection between the bearing support and the sensors. Usually, a piezoelectric accelerometer is externally installed on the machine casing, on the bearing housing, or on a portion of the casing to collect the vibration signals [17]. Among all the sensor types presented in this section, accelerometers are the best tradeoff between reliability, feasibility, and cost.

1.3 Signal process techniques for bearing condition monitoring

Vibration measurement is the most widely used condition monitoring technique for bearings. Ninety percent of machinery failures can be accurately identified by monitoring changes in vibration signals [18]. Therefore, in this work, we focus on the vibration signals.

The bearing degradation mechanisms are complex because usually, one failure mode can initiate another, e.g., an abrasive left by corrosion in a ball race can cause wear which results in loss of preload or an increase in radial clearance [10]. The wear debris can result in lubrication failure and subsequent overheating. Moreover, background noise and electronic devices usually affect the quality of the condition monitoring data. All those lead to the need to process the collected condition monitoring data for mining the information embedded in raw vibration signals.

In order to provide an accurate estimation of the bearings' condition, the raw vibration signals need to be processed to remove all disturbance effects, and to extract features or indicators able to quantify the bearings degradation process efficiently. In the sequel, we propose a brief presentation of existing process techniques for vibration signals.

1.3.1 **Traditional process techniques**

Time-domain process techniques

The most common time-domain features are the statistical time-domain features for the interpretation of the time-domain waveform. As we know, when the condition of the bearing changes, the probability density function of the signal also changes. The statistical time-domain features rely on examining the probability density function of the signal to characterize the condition of bearings.

Common statistical time-domain features are: root mean square (RMS), mean, standard deviation (STD), peak-magnitude-to-RMS level (PMR), kurtosis, and skewness.

The RMS, derived from the vibration signals, provides additional information on the energy quantity of the signal. It is an important feature used to estimate the bearing's degradation [19]. It usually increases gradually as the fault develops. The definition of RMS is provided by the formula (1.1). The mean quantifies the mean amplitude of signals, defined in formula (1.2). Note that, for raw vibration signals collected from bearings degradation, the mean remains close to zero but is not zero [20]. This is due to the inherently dynamic characteristic and structural defects of bearings, environmental noises, and irregular shifts of mechanical structures related to bearing degradation. Vibration signals in rolling bearings are non-stationary in nature[21][22]. The STD measures the dispersion of the signals around the mean value. Its definition is provided by the formula (1.3). The PMR quantifies the pulses' intensity of the signal and it is defined by the formula (1.4). A large PMR value corresponds to a vibration signal with a few large spiky peaks. Similarly, a low PMR value is relevant to a flattish vibration signal [23]. The mean and standard deviation values of the distribution are the first two parameters of the sample distributions. The third parameter of the distribution is skewness and the fourth is the kurtosis. The kurtosis reveals the degree of deviation of the signal distribution from the average and the skewness indicates the signal distribution asymmetry, their definitions are respectively in formula (1.5) and (1.6).

$$RMS_k = \sqrt{\frac{1}{N_k}\sum_{n_k=1}^{N_k} v_{n_k}^2} \tag{1.1}$$

$$mean_k = \frac{1}{N_k}\sum_{n_k=1}^{N_k} v_{n_k} \tag{1.2}$$

$$STD_k = \sqrt{\frac{1}{N_k - 1}\sum_{n_k=1}^{N_k}(v_{n_k} - mean_k)^2} \tag{1.3}$$

$$PMR_k = \frac{\|v_{n_k}\|_\infty}{RMS_k} \tag{1.4}$$

$$kurtosis_k = \frac{1}{N_k}\sum_{n_k=1}^{N_k}\frac{(v_{n_k} - mean_k)^4}{STD_k{}^4} \tag{1.5}$$

$$skewness_k = \frac{1}{N_k}\sum_{n_k=1}^{N_k}\frac{(v_{n_k} - mean_k)^3}{STD_k{}^3} \tag{1.6}$$

where n_k=*1,2,3,…,* N_k is the label of k^{th} sample interval of vibration signals. v_{n_k} is the $n_k{}^{\text{th}}$ value of k^{th} sample interval of vibration signals. More generally, the feature *h*-order moment of the form defined in (1.7) also could be used as the statistical time-domain features.

$$ordermon_k = \frac{1}{N_k}\sum_{n_k=1}^{N_k}\frac{(v_{n_k} - mean_k)^h}{STD_k{}^h} \tag{1.7}$$

where $h \geq 5$, *ordermon* denotes the feature *h*-order moment.

Frequency domain processes techniques

Frequency domain processes techniques aim to identify and isolate different frequency components of useful information from signals. Among these techniques, the Fourier transform is a common and famous technique to transform signals from time-domain into frequency domain [24]. Because of the rolling behaviors, the contact between rolling elements and defective spots of bearings could cause the repetitive impulse period of vibration signals when faults occur. The Fourier transform measures the dominant frequency of vibration signals corresponding to specific faults. Thus, this technique is used to identify different bearing defects under different bearings degradation conditions. The most popular type of Fourier transform is the Fast Fourier Transform. It is effective for stationary signals whose frequency components do not change over time.

Time-frequency domain processes techniques

Time-frequency domain process techniques, such as the Short-Time Fourier Transform and Wigner-Ville distribution, can deal with non-stationary signals.

The Short-Time Fourier Transform aims to describe the frequency characteristics as functions of time through sliding windows of time [25]. The main limitation of the Short-Time Fourier Transform is that there is a trade-off between the resolution of time and the resolution of

frequency. In other words, obtaining a finer frequency resolution must be accompanied by a worse time resolution and vice-versa [26]. The Wigner-Ville distribution, part of the Cohen class of distributions, is a basic time-frequency representation that is sensitive for non-stationary signal analysis [27]. The difficulty of Wigner-Ville distribution is the severe cross terms as indicated by the existence of negative power for some frequency ranges [26].

1.3.2 **Advanced processes**

In recent years, with the advancement of mathematics and other computation technologies, vibration signal processing techniques for bearing monitoring have also made considerable progress.

Spectral kurtosis

The spectral kurtosis, the extended concept of the kurtosis, decomposes the power of a signal based on the Short-Time Fourier Transform and gives a measure of the impulsiveness of a signal as a function of frequency. Spectral kurtosis is a powerful tool to detect the transients in a signal with strong additive noise.

In [28], the spectral kurtosis is obtained from a filter function which filters out that part of the signal with the highest level of impulsiveness where an impulsive bearing fault signal is dominant. The optimum filter is a narrow band filter at the maximum value of spectral kurtosis aiming to maximize the signal-to-noise ratio of the filtered signal without considering its shape. Obtaining the optimal filter for the spectral kurtosis critically depends on the Short-Time Fourier Transform window length, or on the bandwidth of the band-pass filter that outputs the complex envelope. Hence, the spectral kurtosis can also be described as a function of the Short-Time Fourier Transform, giving rise to a two-dimensional representation called kurtogram [28]. The approach proposed in [28] was applied to bearing signals corresponding to an outer race and a ball fault. The numerical results show that mechanical failure signatures have been better extracted compared with other methods.

In [29], a two-step technique is presented to optimize spectral kurtosis and get the kurtogram for diagnosing rolling element bearing faults. First, signals are filtered using an autoregressive filter. Then, the residual signal is analyzed using complex Morlet wavelets, which represents the noise and white spectrum. These are used as a filter bank with constant proportional bandwidth. Different banks are used to select the best filter in terms of center frequency and bandwidth. This method was applied to an acceleration signal corresponding to a bearing with

an inner race fault, running at around 3000 rpm. The numerical results show that the algorithm proposed has a good performance for fault identification by identifying the frequency in the envelope spectrum.

In [30], wavelet packet transform based on the Daubechies wavelet is introduced into kurtogram for improving the accuracy of kurtogram in extracting transient characteristics from a noisy signal and identifying machinery fault. The detection of faults on the outer race of rolling element bearings using vibration signals validates the effectiveness of the proposed method in diagnosis for rolling element bearings.

In [31], for enhancing the surveillance and diagnostic capability of spectral kurtosis for bearings, authors used the autoregressive models based on the linear prediction filtering and minimum entropy deconvolution. The effect of the transmission path is deconvoluted by minimum entropy deconvolution method from the impulses contained in the residual of the autoregressive model. Then, the deconvoluted impulses are decomposed using complex Morlet method. Finally, the spectral kurtosis is calculated for the best filter selection to implement the envelope analysis. Vibration signals of a spall on the inner race and a notch fault on the outer race of a double-row ball bearing at speed around 11760 rpm are respectively used to validate the advantage of the presented algorithm for diagnosis of rolling element bearings.

In [32], to enhance the ability of kurtogram for highlighting the type of fault, kurtosis values are calculated based on the power spectrum of the envelope of the signals. The enhanced kurtogram is an efficient way to locate the resonant frequency bands. The outer race, inner race and rolling element faults of rolling element bearings at speed around 1400 rpm are used to validate the performance of the method for diagnosis of rolling element bearings.

Cyclostationary analysis

Cyclostationarity characterizes a subclass of non-stationary signals that exhibit some cyclical behavior. Due to the inherent periodic modulations, cyclostationarity ideally fits the property of many rotating and reciprocating machine vibrations [33]. Bearings vibration signals verify the capability of the proposed method under conditions of inner race fault, outer race fault, and the cage fault.

Cyclostationary analysis of a vibration signal analyses both the content of a signal and the characterization of periodical content evolution in time to mine useful information.

In [34], to be sensitive to the impulsiveness and the cyclostationarity for rotating machinery diagnostics, a multiscale clustering grey infogram is proposed which involves both time-domain and frequency-domain spectral negentropy in a grey fashion by multiscale clustering.

The capability of the proposed method is verified by bearings vibration signals under conditions of inner race fault, outer race fault and normal condition.

In [35], the authors claim that the squared envelope spectrum is an efficient indicator for the assessment of second-order cyclostationary symptoms of rolling element bearing faults. An analytical derivation of the statistical tests for cyclostationarity in the squared envelope spectrum is proposed without the hypothesis of white noise from the beginning. The test is conducted by vibration signals of a healthy bearing and an inner ring damaged bearing, running at a speed of 2000 rpm and recording data for 10 s, at the sample rate of 10 kHz.

In [36], the slice spectrum analysis of the cyclic bispectrum is proposed to disclose the latent periodic components in the vibration signals with the heavy noise background. The performance of the method is evaluated based on the vibration signals corresponding to defects in the outer race of rolling element bearings.

In [37], subsampling in frequency identification is used for frequency analysis of cyclostationary signals of bearings. The vibration signals of bearings with the fault in the inner race or in rolling elements are used to test the effectiveness of the proposed method.

The research work presented in [38] proposes a bearing diagnosis method based on the autocovariance function of a 2nd order cyclostationary signal. Instead of computing the kurtosis of the filtered time signal, it calculates the kurtosis of the unbiased autocorrelation of the squared envelope of the demodulated signal. Moreover, two modified forms of kurtosis are proposed to take advantage of the unique features of the lower and upper portions of the autocorrelation. The vibration signals of bearings with faults on the inner race are used to test this bearing diagnosis method.

Wavelet dependent techniques

Essentially, wavelet transforms reveal some inner features of the signal, involving a family of wavelets.

Wavelet analysis provides multi-resolution in time-frequency distribution for easier detection of hidden components of vibration signals. Based on the results of extensive experiments, the methods based on wavelet analysis are efficient in detecting bearing faults from abnormal vibration signals [39]. Both periodic and aperiodic signals are suitable for failure detection using wavelet analysis instead of the Fast Fourier Transforms. Vibration signals of the bearings with outer race fault, inner race fault, ball fault and the multi-fault are used to verify the effectiveness of this method for bearings diagnosis.

Wavelet transform is used in [40] for bearing race faults diagnosis. It provides a variable resolution time-frequency distribution from periodic structural ringing due to repetitive force impulses which appear periodically with a time-period corresponding to characteristic defect frequencies. The vibration signals of excellent and defective bearings (faults on the inner race, outer race or balls) have been considered the results section in [40].

An advanced version of the famous "soft-thresholding denoising method" based on the Morlet wavelet is proposed in [41] to extract useful features from vibration signals with the low signal-to-noise ratio. The vibration signal of an inner-race damaged rolling bearing is used to test the effectiveness of the proposed method for bearing diagnosis.

Wavelet analysis is used to process the bearings signals for multi-fault diagnosis in [42]. Vibration signals of defects of the inner race, out race and balls of rolling element bearings, are used to validate the accuracy of the method proposed.

The importance of the appropriate wavelet function selection is emphasized in [43]. In this paper, a Gaussian wavelet-based transform method wavelet is presented for the defect localization in the time domain signal of rotating system vibration.

An improved wavelet package transform is proposed in [44] to extract salient frequency-band features from bearings vibration signals, by constructing a biorthogonal wavelet with impact property via lifting scheme. Data samples with the fault defect size of 0.007 and 0.014 inches under four different loads (0, 1, 2, and 3 hp) are used to verify the capability of the method proposed, which is for classifying bearings faults on the inner race, outer race and balls.

An adaptive morphological update lifting wavelet is used to process bearings signals to extract fault features for bearings diagnosis in [45]. A morphological dilation-erosion filter or an average filter is adaptively used as the update lifting filter to modify the approximation signal. The vibration signal collected from the bearing with localized defect on the outer raceway is used to test the proposed method for bearings diagnosis.

The lifting scheme of wavelet and its multiphase expression is used for the diagnosis of gears, rolling bearings and rotor rubbings in [46]. The numerical results, in this paper, show the effectiveness of the lifting scheme of wavelet for rotary mechanics fault characteristics extraction from the raw signal.

Due to the correlation between structures and the complexity of the equipment, the compound faults are common. This is caused by merging faults in rotating machinery. Single wavelet-based processing methods cannot extract features with multiple kinds of shapes. Therefore, a multiwavelet transform extends the wavelet transform theory to perform a multi-resolution analysis. It also has such important properties of orthogonality, symmetry, compact support and

higher order of vanishing moments [47]. In [48], lifting schemes were used to construct changeable cubic Hermite multiwavelets instead of the fixed ones for mechanical fault detection. The vibration signals of bearings with an incipient fault of a slight rub on the outer race are used to test the method for bearings incipient fault detection.

Multiwavelet with improved neighboring coefficients is used in [49] to extract the features from the bearings' condition monitoring signals in the presence of important noise. Vibration signals of bearings with an outer-race defect are used to test the proposed method for rolling bearing fault diagnosis.

In [50], optimal multiwavelets are used to detect the bearings' faults from non-stationary vibration signals with a large amount of noise. Two-scale similarity transforms are used in the optimal multiwavelets to establish a changeable and adaptive multiwavelet library to be the various ascendant multiple basis functions for inner product operation. Vibration signals of bearings with a slight rub on the outer race are used to test the proposed method for rolling bearing fault diagnosis. The results show that the proposed method is effective for the detection of impulse feature components hidden in vibration signals, and for rotating machinery fault diagnosis.

Based on an adaptive lifting scheme where a genetic algorithm is used to optimize the free parameters, a new multiwavelet is proposed in [51], to analyze a simulation rolling bearing and gearbox vibration signal for rotating machinery fault diagnosis. Vibration signals of bearings with a slight rub on the outer race are used to test the proposed method for rolling bearing fault diagnosis.

To overcome the drawbacks of multiwavelet such as fixed basis function and critically sampled filter-bank, customized redundant multiwavelet is proposed in [52] for rotating machinery diagnosis based on increasing multiplicity. Vibration signals of bearings with faults on the outer race are used to test the proposed method for rolling bearing fault diagnosis.

The wavelet packet transform is a generalization of the wavelet transform, which provides a library of orthonormal basis functions and can, therefore, represent a signal in many different ways [53][54].

The wavelet packet transform is used in [55] to uncover the local features and spectral contents of vibration signals for bearing fault diagnosis. Vibration signals measured for bearings with inner race fault or outer race fault are used to verify the proposed method.

The paving of a wavelet packet transform is used in [56] for rotating machinery diagnosis. Vibration signals are used to evaluate the proposed method for bearings with normal conditions, inner race fault, outer race fault or balls fault.

Wavelet packet decomposition is used in [57] to decompose vibration signals measured into a set of sub-frequency band signals for multi-scale permutation entropy extraction. The results are used to distinguish and diagnose fault feature of rolling bearing under different fault conditions. The method is applied to vibration signals collected from a healthy bearing, a bearing with an inner ring defect, a bearing with a rolling element defect, and a bearing with an outer ring defect.

For signal decomposition analysis, in addition to Wavelet analysis, there is another famous method called empirical mode decomposition, which is described in the next subsection.

Empirical mode decomposition

The empirical mode decomposition method decomposes a signal into some intrinsic mode functions [58]. An intrinsic mode function represents a simple oscillatory mode embedded in the signal. The empirical mode decomposition method shows significant advantages in processing nonlinear and non-stationary signals.

The approach presented in [59] uses the empirical mode decomposition to characterize the bearing condition based on an energy entropy feature that changes to different frequency bands when bearing fault occurs. The performance of the proposed method was verified using vibration signals collected on bearings with normal conditions, inner race fault, or outer race faults.

In [60], the empirical mode decomposition method is used to process vibration signals to extract fault characteristic information. Vibration signals of normal condition, outer race fault, inner race fault, and ball fault) with the fault defect size of 0.007, 0.014, and 0.021 inches under four various loads (0, 1, 2 and 3 hp) are used to verify the capability of the method proposed for fault classification of bearings.

In [61], the empirical mode decomposition method is used to decompose the original modulation signals into several intrinsic mode functions. The bearing condition is assessed based on characteristic amplitude ratios defined as the ratios of amplitudes in the envelope spectra of intrinsic mode functions at the different fault characteristic frequencies. The proposed method was validated using the vibration signals of bearings with normal conditions, inner race fault or outer race faults.

In [62], based on vibration signals, the empirical mode decomposition method is used to extract features to form a health index to characterize the bearing conditions. The lifecycle vibration signals of bearings ended with roller failure, inner race failure, or outer race failure are used to verify the proposed method.

In [63], based on vibration signals the empirical mode decomposition method, energy entropies are extracted from vibration signals to characterize the bearing conditions. Vibration signals of bearings are used to test the proposed method in four different operating conditions (normal condition, outer race fault, inner race fault, and ball fault) and four different loads (0, 1, 2 and 3 hp). Moreover, each fault condition includes two different defect sizes of 0.007 and 0.021 inch, respectively. In [64], the empirical mode decomposition method is used to obtain local Hilbert marginal spectrum with Hilbert transform method, based on an envelope signal extracted from vibration signals by envelope spectrum analysis of wavelet coefficients of high scales. The diagnosis of bearings in this work is based on the obtained local Hilbert marginal spectrum. Vibration signals of rolling bearings with outer race fault or inner race fault are used to test the proposed method. Ensemble empirical mode decomposition proposed in [65] is an improved version of empirical mode decomposition. Its advantage is that it reduces the occurrence of mode mixing comparing with the empirical mode decomposition.

In [66], ensemble empirical mode decomposition is used to decompose the vibration signal into several intrinsic mode functions. Permutation entropy values of the first several intrinsic mode functions are calculated to form the fault feature vector. To assess its effectiveness for fault classification, the method proposed was applied to classify the expected condition, outer raceway faults, inner raceway faults and balls faults of bearings based on the vibration signals derived from the drive end bearing data collected at 12,000 samples/s under four different loads, 0–3 hp with fault diameters 0.007–0.028 inch.

In [67], ensemble empirical mode decomposition is used to extract energy entropy and singular values of the matrix whose rows are intrinsic mode functions. Energy entropy is used to detect the existence of faults. Singular values are used to identify the nature of the fault. Vibration signals of bearings with faults on the inner race, outer race, and ball with diameters 0.007 inch and 0.021 inch collected at four different loads of 0–3 hp (at motor speeds of 1720–1797 rpm) are used to test the proposed method.

Hybrid methods

The previous content shows that most of the single signal processing techniques are only effective for uncovering the signature in the signals of some specific kinds of bearings condition cases. On the one hand, there are heavy noises in bearings condition monitoring signals. On the other hand, correlations between the degradation process and different information carried by bearings condition monitoring signals are diverse. Compared with methods based on single signal processing techniques, it is natural that hybrid methods have advantages for eliminating

the noises and extracting useful information from signals. We classify the hybrid methods into the following categories according to techniques implemented.

a) Denoising and then analyzing category

In [68], wavelet packet decomposition is applied to denoise signals, then the ensemble empirical mode decomposition technique is used to extract informative feature vectors. Vibration signals of bearing with a damaged roller are used to verify the proposed method. A method combining morphological filter with translation invariant wavelet is used in [69] to reduce the narrowband impulses and random noises in the original signal. Then the purified signal is decomposed using improved ensemble empirical mode decomposition to obtain intrinsic mode functions. Finally, the envelope analysis is used to extract defect information form selected intrinsic mode functions. Vibration signals of defective bearings respectively with an outer race fault, an inner race fault and a rolling element fault are used to test the proposed method.

The difference between the closing and opening operator is employed in [70] as the morphology filter to extract the impulsive periodicity features from the purified signal which is obtained from the raw signal by the second-generation wavelet. Vibration signals of defective bearings respectively with an inner race fault, outer race fault, a rolling element fault and the compound faults are used to test the proposed method.

In [71], first, through the signal-to-noise ratio of probabilistic principal component analysis denoising model, noises and linear interrelated information of signals will be discarded, based on projection to the residual subspace. Based on the denoising signals, a band-pass filter is used to determine optimal center frequency and bandwidth. Finally, a Hilbert envelope spectrum analysis of the filtered signal is performed to extract the fault frequencies of the rolling element bearings. Vibration signals of bearings, respectively, with the inner race fault and rolling element fault, are used to evaluate the proposed method.

The second-generation wavelet transforms using neighboring coefficients is used in [72] to remove noise in rotating machinery vibration signals. Second generation wavelet transform has the merit of enhancing the signal-to-noise ratio. Then, the local mean decomposition method is used to decompose the de-noised signals into product functions corresponding to the faulty feature signal selected according to the correlation coefficients criterion. Finally, the Fast Fourier Transform is used to analyze the frequency spectrum. Vibration signals of a bearing with localized outer-race defects are used to verify the proposed method.

b) Analyzing and then feature selection category

In [73], the extraction process is based on trigonometric functions and cumulative transformation. The feature selection is performed by evaluating feature fitness using monotonicity and trend ability characteristics. Vibration signals of bearings over completely run to failure life cycle under 3 different load conditions are used to verify the proposed method.
In [74], first, empirical mode decomposition is used to decompose a vibration signal into intrinsic mode functions. Then, a modified genetic algorithm based on variable-range encoding and dynamic searching strategy is used to establish relationships between optimized feature subsets and the classification performance. Finally, the dominant features are selected for characterizing the bearing conditions is executed by a statistical model based on receiver operating characteristic. Vibration signals of bearings with inner race faults, outer race faults, and rolling element faults of three different fault sizes are used to test the proposed method.

c) Analyzing and then dimensionality reduction category

The vibration bi-spectrum (third-order spectrum) patterns are extracted to characterize different types of bearing faults in [75]. Then, the dimension reduction is implemented using the principal component analysis method. Vibration signals derived from bearings under conditions of healthy, outer race fault, ball fault and inner race fault were used to verify the method proposed for bearings fault classification.
The vibration signals are used in [76] to build a high-dimensional feature made of statistical features, autoregressive coefficients, and instantaneous amplitude Shannon entropy from the intrinsic model functions of empirical mode decomposition. For obtaining more sensitive low-dimensional fault features, it is projected into an orthogonal supervised linear local tangent space alignment for dimension reduction. The vibration signals of bearings with normal state, roller-element fault, inner-race fault, and outer-race fault at a speed of 2100 rpm (35 Hz) and 10 kHz sampling frequency are used to test the proposed method.
In [77], first, the energy entropy of intrinsic mode functions are obtained by the empirical mode decomposition. Then principal component analysis and linear discriminant analysis are used for feature reduction. Vibration signals of 12 bearings are used to test the proposed method, among which only 4 bearings reached the failures respectively caused by defects on the inner race, outer race, and rolling elements.
In [78], first, the intrinsic manifold features from high-dimensional feature vectors are extracted respectively from time-domain, frequency-domain, and empirical mode decomposition based on vibration signals. Then, to reduce feature dimensionality, a statistical locally linear embedding algorithm is used to translate the complex mode space into a salient low-

dimensional feature space. Vibration signals of bearings with roller-element fault, inner-race fault, and outer-race fault are used to verify the proposed method.

In [79], first, the Shannon mutual information between all samples and training samples is combined to represent the mutual dependence of samples, which is the incipient fault features. Then, a method called Supervised Orthogonal Local Fisher Discriminant Analysis is proposed, which aims to compress fault features (the high-dimensional Shannon mutual information sets) of testing and training samples into eigenvectors with low dimensionality, which have clearer clustering. Vibration signals of bearings with normal state, tiny outer race cracks, tiny outer race cracks, and tiny ball cracks are used to verify the proposed method.

In [80], first, a set of original features from time, frequency, and time-frequency domain are extracted. Then, the neighborhood component analysis based feature extraction approach is proposed to reduce the dimensionality of the original features set. Vibration signals of bearings with normal state, roller-element fault, inner-race fault, and outer-race fault are used to test the proposed method for fault type classification. The vibration signals collected over the whole life of bearings are used to verify the proposed method for defect severity classification.

In [81], first, mechanical fault vibration signals are decomposed into multiple intrinsic mode functions using variational mode decomposition. Then time-frequency domain features are extracted from intrinsic mode functions to construct the feature sets. Finally, semi-supervised locally linear embedding is used for fusion and dimension reduction. Vibration signals of bearings with normal state, roller-element fault, inner-race fault, and outer-race fault are used to test the proposed method for bearing diagnosis.

1.4 Bearings remaining useful life estimation

1.4.1 Bearings useful life

Bearings useful life is defined as the time duration between the start of the operation and the first moment of failure, as depicted in Figure 1.13.

When repeated stresses are applied to the material with finite volume in bearings, failure can occur in weak places in the material. In general, on the microscale, the material does is not have a homogeneous structure. Therefore, it is usually characterized by a large dispersion in strength and resistance. Different parts of bearings suffer from repeated stresses in different ways and different rates. All these properties result in that there are significant variations on bearing useful life in either laboratory tests or practical experience, even for identical bearings under identical load and environmental conditions [82].

Usually, in industrial applications, the usual bearing useful life metrics are L10 life and L50 life. The L10 life describes the operation duration until failure occurs to 10% of a large enough sample of identical bearings under identical conditions. Similarly, the L50 life defines the operation duration until failure occurs to 50% of a large enough sample of identical bearings under identical conditions.

However, L10 life and L50 life are theoretical concepts based on statistical analysis. For a specific bearing in real industrial application, it is not guaranteed that, for instance, the bearing has a 90% reliability after the amount of time of L10 life has passed [83].

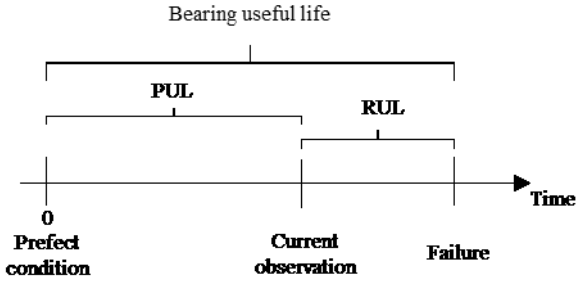


Figure 1.13 Bearing useful life.

There are two important concepts related to bearing useful life: the past useful life (PUL) and the remaining useful life (RUL), as shown in Figure 1.13.

Because of the significant variations on bearings' useful life and the crucial roles of bearings in the mechanical systems, the issue of bearings RUL estimation has attracted much attention from engineers and scholars.

1.4.2 **Methods for RUL estimation**

The event data based approaches use event data of a population of identical objects to assess the reliability based on the distribution analysis and estimate the RUL. Several parametric failure models can be adapted to estimate the reliability of rolling machinery elements. For instance, the exponential models work well for the RUL prediction with exponential-like degradation processes [84]. The Poisson based models could capture the stochastic characteristics of the degradation evolution for bearings reliability analysis. The Weibull distribution is often used to estimate the bearings' probability of failure at a specific observation, mean lifetime and failure rate. Log-Normal distribution based models have the advantage to model the stock behavior of the degradation evolution for bearings reliability

analysis [85]. The event data based approaches aim to characterize the entire population of identical objects rather than individuals [86].

The development of condition monitoring technologies allowed collecting condition data on individual bearings. The condition data provide valuable information associated with the health condition of individual bearings, and several approaches use this information for fault diagnosis and RU assessment. Meanwhile, some approaches could integrate event data with condition data to provide the estimated RUL as well as the statistical characteristics of RUL.

The bearing degradation mechanism is complex and it is not easy to build an accurate mathematical model. There are mainly three categories of approaches for estimating the remaining useful life of bearings: failure physical model approaches, statistical data-driven approaches and artificial intelligence approaches [86].

Failure physical model approaches

Failure physical model approaches, e.g., Paris crack growth model, Forman crack growth model, spall progression model, and so on, describe the degradation process according to the variations in the physical states and evolution of structural damage. They follow the mechanical laws governing the material response to loading conditions in bearings.

A Paris's formula based deterministic defect propagation model is proposed in [87] to describe the diverse defect propagation in bearings. To fit the dynamic degradation process of bearings, an adaptive algorithm based on non-linear recursive least square is used to update the parameters of the model recursively. Considering the time-variant nature of defect growth, the proposed method outperforms the deterministic models without adaptation in both continuous and discontinuous defect growth cases.

A multi-time scale modeling approach is proposed to estimate bearings RUL in [88]. The Paris crack growth model is used to describe the degradation process in a slow-time scale. On a fast-time scale, the degradation process is described by a phase space warping technique. The proposed method is applied on two bearings run to failure test to show the multi-time scale modeling approach could improve the RUL prediction.

An integrated prognostic approach is proposed in [89] to estimate bearings RUL, based on Paris's law and a Kalman smoother. The kurtosis is used as a health indicator. Its trend increases with large fluctuations along with the evolution of bearings degradation.

In [90], a Forman crack growth law of linear elastic fracture mechanics based life model is proposed to estimate RUL.

In [91], an integrated prognostic approach is proposed. The spall progression model is used to estimate bearings RUL at the early stage of the degradation process.

The RUL estimation approaches based on the physical failure model often assume that the degradation process can be described by a single or multiple physical models, which relies on the physics of the underlying degradation process to predict the onset of failures. However, in practice, it is not easy to build accurate physical models for RUL prediction in case of complicated failure mechanisms.

Statistical data-driven approaches

a) Stochastic models

Statistical data-driven approaches rely on stochastic models. Gamma process model, inverse Gaussian process model and Wiener process model are usually considered [92].

A degradation evolution, caused by wear and cumulative damage, shows a monotonic trend typically. The Gamma process is the limit of a compound Poisson process which is monotonic. There is an assumption that external shocks could cause degradation phenomena. Moreover, a compound Poisson process describes the shock arrival process well. Hence, the Gamma process could be a good choice for the degradation process [93]. A Gamma process-based RUL estimation method is mentioned in [94]. The authors assume a monotonic evolution of the degradation process. The parameters of the model are estimated from condition data by the maximum likelihood estimator.

In [95], the physical meanings of the inverse Gaussian process are discussed. The inverse Gaussian process is also a limit of compound Poisson processes as the Gamma process. For the inverse Gaussian process, covariates and random effects can be easily embedded in order to model the degradation caused by random environments.

In a framework of infinite time, the increment of degradation evolution can be considered to be an additive accumulation of numerous tiny external effects. According to the central limit theorem, the additive accumulation is approximated by a normal distribution. Since these tiny external effects and the accumulation in disjoint durations typically are independent, the Wiener process is suitable for modeling the degradation evolution whose increment respects the normal distribution [93].

In [96], bearings continuous degradation and randomly arriving shocks are both considered. A Wiener process with a linear drift is used to model the degradation process. For the randomly arriving shocks, a compound Poisson process is adopted to model the impacts of the randomly

arriving shocks for the degradation process. Moreover, the probability density function of the RUL is estimated, and it is updated along with the condition data.

In [97], authors assume that systems show different deteriorating rates in the different operation states, which considers external stresses and mutation in the performance. Considering this assumption, a nonlinear Wiener process with state switching depicted by a continuous-time Markov chain is used to describe the operation process.

A hybrid model is proposed in [98] for RUL estimation. A nonlinear Wiener process models the continuous degradation process, while a nonhomogeneous compound Poisson process models the impact of randomly arriving shocks. The combination of the expectation conditional maximization algorithm and maximum likelihood estimation is used to update the parameters of the model along with the degradation evolution.

Wiener process is suitable to model a non-monotonic degradation process with Gaussian noise. In [99], the continuous degradation process and randomly arriving shock are also considered. For RUL estimation, the Wiener process is used to model the continuous degradation process, while a Poisson process is used to describe the random shocks arrive.

In [100], the condition data are decomposed into trend items and random items, which are respectively associated with the continuous degradation process and randomly arriving shock. The Wiener process and the normal stochastic process are used to model the continuous degradation process respectively and randomly arriving shocks for RUL estimation.

In [101], a Wiener process-based degradation model for real-time bearings RUL estimation is established according to the degradation evolution of bearings. The parameters of the model are set using the maximum likelihood estimation method.

In [102], the authors assume that there are two different degradation stages, the normal stage, and the degradation stage, before failure. An improved Wiener process-based model is established for the second stage, which also describes the uncertainties of the degradation evolution. The parameters of the nonlinear model are considered as hidden states in state-space modeling and they are updated simultaneously.

In [103], the authors assume that two different degradation stages before failure are mutually dependent. A two-stage Wiener process model is proposed, which involves stage correlation to the prior distributions of model parameters.

Parallel simulation for RUL estimation is an emerging simulation technology. Using real-time condition data, model structures, and model parameters are adapted. In practice, the equipment with complex structure normally has a mixture of degradation evolution with discrete shocks. So, it is necessary to adapt both the model structure and the model parameters along with the

evolution of degradation. In [104], a multiple model filtering algorithm is used along with the degradation evolution to adaptively switch between two types of models respectively based on the Wiener process and Poisson effect. For estimating the model parameters, the expectation-maximization algorithm is used along with the degradation evolution. Iteratively, the actual equipment degradation states are accurately approximated.

Stochastic models such as the Gamma process model and the Wiener process model have the advantages of relatively straightforward mathematical calculations and the easy understanding of physical meaning. However, they are suitable only for modeling strictly monotonic process such as wear processes or fatigue crack propagation without considering other characteristics of bearing degradation evolution.

b) Markov and Bayesian-based methods

There is a strong mathematical basis in the Markov and Bayesian theory for reliability analysis and RUL estimation.

In [105], authors assume that there are three stages of bearing degradation evolution, i.e., normal operation, defect initiation, and accelerated performance degradation. This paper presents a multi-mode particle filter to automatically detect the transition among three degradation stages and accurately characterize bearing performance degradation by stochastic modeling method. These modes are updated along with the bearing degradation evolution. A finite-state Markov chain is used to switch among these modes, which reflects the transition among the degradation stages.

In [106], the authors assume that there are two phases in the bearing degradation process. A two-phase threshold model is used to model different phases of degradation. Using Bayesian methods, the posterior distributions of the model parameters are updated along with the degradation evolution. Also, the proposed method considers the correlation among the multiple predictions to improve the accuracy of RUL distribution.

In [107], a switching Kalman filter is used for RUL estimation, which is a state-space based method. Bayesian estimation is used to assist in selecting multiple models according to the probability of every state at each time step. This approach can provide a probabilistic measure of the bearing degradation process and predict in the case of the degradation is unstable. The proposed method assumes that the degradation rate is increasing monotonically.

An online remaining useful life prediction method is proposed in [108] to recognize the linear degradation patterns under the framework of a generalized nonlinear degradation model with deterministic and stochastic parameters. Using the maximum-likelihood estimation method, the

prior distribution of the stochastic parameters is estimated. The Bayesian paradigm is used to update stochastic parameters in the degradation model along with the degradation evolution.

Multiple degradation features extracted from vibration signals of bearings were selected and fused as the health index using dynamic principal component analysis and Mahalanobis distance methods [109]. Based on the health index extracted, an exponential regression-based local degradation model is used to model the local degradation process. An empirical Bayesian algorithm is used to update the local models along with the degradation evolution to model the global degradation process for bearings RUL estimation.

A two-stage Wiener process model is used in [110] to model the degradation process with stage correlation. The authors assume that the model parameters in different stages are mutually dependent. A Bayesian approach is used to involve the stage correlation into the prior distributions of model parameters to embed more information.

Hidden Markov model is a statistical data-driven approach for bearings RUL estimation. Briefly speaking, considering the observations as discrete symbols, it uses discrete probability densities to model the transition and the observation probabilities of different stages in bearings degradation evolution [111].

Dynamic Bayesian Networks are used in [112] to describe a Mixture of Gaussians Hidden Markov Models to model the degradation for RUL estimation of bearings operating at constant load conditions and constant speed.

Mixture of Gaussians Hidden Markov Models is established by dedicated learning algorithms in [111] from features extracted by Wavelet Packet Decomposition to estimate the RUL of bearings operating at constant load condition and constant speed.

In [113], a health index is extracted by the Hidden Markov model through the principal features obtained by the principal component analysis method. Then, an adaptive stochastic fault prediction model is developed for on-line bearing RUL prediction.

In [114], the degradation model of the Multi-Branch Hidden Semi-Markov Model is proposed which has several different branches. The models describe the degradation process in different stages and consider the correlations between different stages. Hence, the model proposed can assess both the current health state and the degradation mechanism. It was applied to model degradation mechanisms characterized by multiple competing modes.

Markov model-based methods keep the assumption that the bearing degradation could be divided into several meaningful states which is similar to common sense in practical situations. However, in this assumption, the future degradation state depends only on the current one. This

memoryless assumption leads to an approximation for the true bearing degradation process. Moreover, the conditional independence between the different states is not always satisfied.

c) Kalman filter

Kalman filter is a recursive algorithm, which predicts the future unknown variables with the joint probability distribution, based on time series of data involving statistical noise and other inaccuracies [115].

In [116], 14 different time-domain features are extracted from vibration signals of bearings. A Kalman filter-based two-stage strategy method is proposed to predict the bearings RUL. The first stage is monitoring the bearings' condition until a degradation point is detected. In the second stage, the Kalman filter-based model is used to estimate the remaining useful life of the bearings. The proposed method keeps the assumption that there are two stages in the degradation process of bearings.

In [117], the Kalman filter is used to associate simulation outputs with observation data for adapting the simulation model along with the degradation evolution. The aim is to use the dynamic evolution of the simulation model in order to provide accurate bearings RUL estimation results. The proposed method assumes that the bearings degradation process is linear.

A model based on extended-Kalman-filtering is established in [115] to estimate the bearings RUL, based on learning the trend of degradation evolution by extracted features derived from the training data.

In [118], a health index is built to describe the degradation evolution based on the condition data of bearings using an autoregressive model. A nonlinear model is established based on an extended Kalman filter to track degradation evolution for bearings RUL estimation.

In [119], a switching Kalman filter based method is used both for bearings diagnostic and prognostic. Multiple dynamical models are used to describe different degradation states. Bayesian estimation is used to infer the most probable underlying degradation states. The proposed method assumes that the degradation rate is monotonic increasing.

In [120], a Switching Unscented Kalman Filter method is proposed for bearing RUL estimation even in case of the condition monitoring data shows a declining trend along with the bearing degradation. The unscented transform is used to avoid errors in the degradation process evolution modeling caused by the property of the linearization in the Extended Kalman filter method. Three kinds of degradation state model changes in the whole bearing life-cycle, i.e., normal running stage, slow degradation stage, and accelerated degradation stage.

Kalman filter, as we know, it is only suitable for a linear degradation process with Gaussian noise. The extended Kalman filter aims to model the nonlinear degradation process. For the nonlinear process, there is a method called Particle filter which described in the next subsection.

d) Particle filter

The particle filter is a recursive algorithm based on the sequential Monte Carlo method and Bayesian inference, which is suitable to describe the nonlinear characteristics of bearings degradation [121].

In [121], the authors claim that the main drawback of the traditional Particle filter is the degeneracy problem. After some recursive steps, the calculation effort to update the useless particles for the posterior distribution is essential. There are two main effective ways to solve the degeneracy problem of the traditional particle filter: the resampling smoothing and the importance density function selection. The particles with large weights are retained in the calculation process through the resampling smoothing to reduce the calculation load. The importance density function aims to reduce errors in subsequent iterations. Therefore, in this paper, a particle filter enhanced with an adaptive importance function selection method and a backpropagation neural network-based resampling smoothing method is used to estimate the bearings RUL. Vibration signals from bearings test-to-failure experiments are used to assess the performance of the method.

A method based on the combination of deep belief network and particle filter is proposed in [122] to estimate bearings RUL. The feature is extracted from raw vibration signals by using the fast Fourier transform. Furthermore, the strength of deep learning is used through deep belief networks to automatically process massive data and accurately predict RUL of the bearings using the particle filter method.

In [123], RMS and peak values calculated from the raw vibration signal are used to monitor the degradation process of bearings. A genetic algorithm is employed to replace the traditional resampling operator for easing the particle leanness problem. A time-varying autoregressive model is used with the Akaike Information Criterion to establish a dynamic model based on Particle Filter. For selecting the starting prediction time, a method proposed in this article is based on hypothesis testing theory.

In [124], a Particle Swarm Optimization algorithm is used to extract the features with an apparent quality trend. Based on the extracted features, the particle filter is used to establish a degradation model, including a nonlinear state evolution and a Gaussian process noise. A Neuro-Fuzzy System is used for RUL estimation.

In [125], health indicators are extracted based on the Hilbert–Huang transform. Then Spearman's coefficient is used to select indicators. Based on the extracted features, a model based on the particle filter method is used to describe the degradation evolution for bearings RUL estimation. A modified recursive least-squares method is proposed to update the parameters effectively in the nonlinear case. A new parameter fusion method based on normalized partial derivative weights is introduced to solve the problem of fluctuation of the model parameters after convergence. A modified maximum likelihood estimation method is proposed to estimate the noise value converging during the training process recursively.
In [126], a degradation rate tracking-based particle filter is proposed to overcome the problems of importance functions selection and the degeneracy problem in traditional particle filter as well as the signal with heavy noise for bearings RUL estimation.
Generally, for the traditional Particle filter, the efforts of authors are focusing on overcoming the drawback of it, e.g., degeneracy problem.

Artificial intelligence approaches

a) Support vector regression

Support vector machines, originally derived from the statistical learning theory, was developed as a general framework to estimate dependencies between finite samples.
A probabilistic support vector regression-based prognostic approach is proposed in [127] for probabilistic RUL prediction. Multiple statistical features from the time-domain, frequency domain, and time-scale domain through a wavelet transform are extracted from vibration signals of bearing. The model is trained and calibrated off-line. Then, the RUL is estimated on-line by the trained model.
In [128], the health indicators are extracted from bearings vibration signals by Hilbert-Huang transform. The results of the diagnosis derived from the support vector machine method are used as the extra health indicator. Based on these indicators, the bearings RUL is estimated using a time-series prediction model by the support vector regression method.
In [129], the isometric feature mapping method is used to perform nonlinear feature reduction to extract the bearings degradation features with a low dimension. The support vector regression model is trained in the off-line phase and applied in the on-line phase for bearings RUL estimation.
A method based on Streaming Data Analysis and the Online Support Vector Regression is proposed in [130] for bearings RUL estimation. This method aims to optimize the trade-off

between the accuracy of the Online Support Vector Regression models and the computational load.

In [131], the bearing degradation process is divided into several stages by a classification model. Local regression models are built from these stages. Support vector machine methods are used to implement both health state assessment and local RUL prediction. The proposed method assumes that a specific regression model may not be able to represent the entire history of the bearings' degradation.

In [132], the least squares support vector regression is used with the hidden Markov model to build the bearings RUL estimation approach. The least-squares support vector regression method is used to predict the trends of features extracted from vibration signals by a linear predictive method. Hidden Markov models are trained to describe different stages of the degradation process. Based on updated data, the model from the least-squares support vector regression is updated to estimate RUL. Furthermore, the probabilities of the predicted features for each hidden Markov model are calculated based on forward or backward algorithms. Future health states and the RUL could be obtained according to those probabilities.

Vibration signals of bearings with inner race fault are used to assess the performance of this method.

In [133], time-domain and frequency-domain features extracted from bearing vibration signals are used to elaborate a degradation indicator based on the self-organizing map. Then a model based on support vector regression prediction is trained to estimate the bearings RUL using the degradation indicator as input and the normalized past life percentage of the bearing as output. The proposed method has the capability of predicting the bearing damage at an early stage.

In [134], a new health indicator is extracted from bearing vibrations signals based on ensemble empirical mode decomposition, Gaussian mixture models, and Jensen-Rényi divergence to overcome the insensitiveness of existed health indicators to incipient defects and a highly fluctuating behavior with the increase in degradation severity. The new health indicator is used to train the support vector regression model. Particle swarm optimization is used to select the hyperparameters selection. Thus, the optimized support vector regression model is used to predict the new health indicator and to estimate bearings RUL with a predefined failure threshold.

In [135], the Empirical Mode Decomposition method is used to reconstruct the signals from raw bearing vibration signals by using a defined criterion. The support vector regression-based model is used to estimate bearings RUL by using the Kurtosis and RMS values of the reconstructed signals.

Support vector machines, introduced by Vapnik [136], are top-rated over the last decade for solving both classification and regression problems. The method can obtain the global optimal solutions without overfitting problem. However, the main drawback of support vector machines is the choice of parameters and kernel function. The parameters setting have an important influence on the performance of the method.

b) Artificial neural networks

The artificial neural network is an adaptive system of a network of artificial neurons that are interconnected. Essentially, it uses mathematical or computational models for processing information. It can change its parameters based on information transformation through the network [137].

A new kind of indicator called minimum quantization error is derived using a self-organizing map, from six vibration features in order to assess the degradation of bearings [138]. This indicator aims to capture the physical transitions that the bearing undergoes during different stages of its life. Then a backpropagation neural network is used to estimate the bearings RUL based on the new indicator.

In [139], the most sensitive features selected from an original feature set using monotonicity and correlation metrics. The original features set is constituted by 11 time-domain features, 5 frequency spectra and 8 time-frequency domain features. Then based on the selected features, a health index is elaborated using recurrent neural networks to estimate the bearings RUL with failure thresholds.

In [140], a new artificial neural network model is proposed for bearing RUL estimation where the inputs are the age and multiple condition monitoring features extracted at the present and previous inspection dates. The past life percentage is the output of the model. All condition monitoring features are previously fitted by a generalized function of the Weibull failure rate to eliminate the influence of random shocks that exist in the bearing degradation process.

A model based on artificial neural networks is also elaborated in [141] to estimate the bearings RUL using time and fitted features corresponding to present and previous time dates. The Weibull hazard rate function is used to fit the RMS and kurtosis features derived from bearings vibration signals to eliminate their fluctuations in the degradation process. The normalized past life percentage is used as the output of the model.

A deep neural network model is proposed in [142], to take advantage of deep learning methods for processing the large-scale data to estimate the RUL of bearings under identical operating conditions. This method relies on three time-domain features and a novel frequency-domain

feature. The time-domain features are sensitive to the middle stage of bearing degradation process, while the frequency domain feature is sensitive to the earlier stage and the later stage.
A deep convolution neural network model is proposed in [143] to estimate the bearings RUL with a large amount of high data dimension, high interference noise and complicated mapping relationship. A new feature named the spectrum-principal-energy-vector is extracted and used as the input of the model. Besides, a smoothing method is used to deal with the discontinuity in the prediction results.
In [144], a nonlinear autoregressive neural network is proposed to estimate the past life percentage of the bearing. To overcome the non-monotonicity of common degradation features, a new health indicator is built based on the extracted features from bearings vibration signals using the Mahalanobis distance criterion and cumulative sum chart.
A long short-term memory recurrent neural network-based model is proposed in [145] to estimate the bearings RUL. The particle swarm optimization method is used to optimize the network structure and the parameters' setting simultaneously. A combination of a new degradation feature called waveform entropy and common degradation features is the input of the model. An automatic detection method is proposed to identify the fault occurrence. According to the detection of fault occurrence, the fault propagation will be tracked and assessed by the model. According to one category of the outputs of this model, a nonlinear exponential degradation model is used for failure time estimation.
A deep end-to-end framework based on convolutional and long short-term memory recurrent units is proposed in [146] to encode the degradation information automatically. The convolutional layer extracts local features. Then, the long-short-term memory layer is used to recognize the degradation process. Using the outputs of the long-short-term memory layer, the bearings RUL is estimated with a failure threshold.
In [147], a multiscale convolutional neural network-based approach is proposed to estimate bearings RUL based on the features extracted by time-frequency representation and bilinear interpolation methods from bearings vibration signals. Comparing with traditional convolutional neural networks, this approach has the advantage of considering the global and local information synchronously by the multiscale layer.
In [148], an increasing monotonous function, globally nonlinear and locally linearizable, is used to model the degradation process. A neural network is used to learn this model and then predict future values of RMS of bearing vibration signals in a run-to-failure test to estimate the RUL of the bearing.

In [149], an artificial neural network-based method is proposed to estimate the RUL. Both the age and multiple condition monitoring data of bearing vibration signals at present and previous time points are as the inputs, and the life percentage is as the output. The input data are fitted by a function generalized from the Weibull failure rate function, reducing the effects of the noise factors that are irrelevant to the bearing degradation in the condition monitoring data.

In [150], to overcome the limitation of data sample size and improve RUL prediction accuracy, an artificial neural network-based multi-step ahead prediction model is proposed for bearing RUL estimation. Predicted features are post-processed and regarded as inputs in the next prediction iteration. Degradation features are divided into adaptive time windows which are adjusted according to an increasing rate. Features in two adjacent windows are regarded respectively as the inputs and outputs of the training set for the multi-step ahead prediction model.

In [151], inspired by the basic form of extreme learning machine, a multilayer neural network with a special structure is proposed to implement the health quantitative modeling and assessment for bearings RUL estimation. It maintains the powerful nonlinear function fitting ability of extreme learning machine while being able to learn the network parameters from qualitative pairwise comparison training samples by a proposed method called pairwise comparison learning.

In [152], an approach based on deep feature representation and long short-term memory neural network is proposed for bearings RUL prediction. To capture the temporal information on fault feature in the non-stationary vibration signals of bearing degradation, the convolutional neural network is used with the Hilbert–Huang transform method, to obtain the deep features of bearing fault with good representation ability. Considering the temporal information of the degradation process, a long short-term memory neural network-based model is used to predict the bearings RUL.

In [153], the generative adversarial networks based model is used to model the degradation process to predict the future trend of degradation indicators for the bearings RUL estimation. The RMS of the bearing vibration acceleration signals from the run-to-failure test as the health indicator of bearing degradation. The indicator RMS is used to train the model in this paper.

In [154], recurrent neural networks-based models are used to estimate the bearings RUL. A recurrent neural network with long short-term memory is used to build a degradation indicator based on the multiple features extracted from bearing vibration signals. Then, based on the extracted degradation indicator, another recurrent neural networks with long short-term

memory is used to predict the trend of the extracted degradation indicator and estimate the bearings RUL.

In [155], to comprehensively assess the bearings degradation process, a comprehensive characterization indicator is obtained from 35 features in time-domain, frequency-domain, and time-frequency-domain by using principal component analysis. Based on this indicator, an artificial neural network is proposed to predict the bearings RUL.

The artificial neural network is an adaptive system of networks of artificial neurons. Essentially, it uses mathematical or computational models for processing the data. It has the ability to change its parameters based on information processing through the network [137].

Artificial neural network-based methods do not use an analytical model of the damage propagation but aim at modeling the damage propagation process, or degradation process, based on the collected condition monitoring data. They are suitable for modeling the bearing degradation evolution without following an established bearing diagnostic or prognostic framework. The main drawback is that a significant amount of data is required for training the model to be accurate.

However, the amount and quality of data collected in practice do not always meet the requirements of artificial neural network training.

c) Fuzzy logic

The fuzzy logic theory is natural for its ability to handle problems with imprecise and incomplete data and model nonlinear functions, which is suitable to describe the nonlinear complex relationship and ambiguous characteristics in the bearings degradation process. Moreover, it has demonstrated its efficiency for the cases with incomplete information [156]. Hence, the fuzzy logic-based method might be suitable for bearings RUL estimation in case of a small size available dataset.

An approach based on simplified fuzzy adaptive resonance theory map neural network and Weibull distribution is proposed in [157] for bearings RUL estimation. The aim is to describe more accurately bearing degradations and to assimilate the prognostic task to a classification one. During the training phase, the Weibull distribution is used to fit degradation features in order to eliminate their fluctuations in the time domain. Using fitted degradation features, a model based on a simplified Fuzzy Adaptive Resonance Theory Map neural network is trained for bearings RUL estimation, which is a fast incremental supervised learning system for analog inputs. Benefit from the fuzzy learning process, the model shows good performance in a test phase with original degradation features.

Fuzzy logic is natural to process linguistic and ambiguous information, which is convenient for modeling the ambiguous aspect of the bearing degradation evolution. To take advantage of fuzzy representation, a hybrid approach based on neural networks and fuzzy logic, called the Fuzzy Back Propagation network, consists of the input layer, hidden layer, and output layer. The nodes, in the hidden layer and output layer, are fuzzy neurons, used in [158] to estimate the bearings RUL.

Fuzzy membership functions and fuzzy rules in adaptive neuro-fuzzy inference systems (ANFIS) are derived via the learning of abundant historical data, which is an efficient way of handling the regression problem. In [159], to take advantage of ANFIS, the adaptive network-based fuzzy inference system is used to predict the bearings RUL, using a fused degradation indicator extracted by globally linear embedding based on various features of bearing vibration signals.

Usually, several features are needed to support accurate monitoring of the bearing degradation. However, multiple features processing needs more calculation effort. Hence, a fuzzy fusion method is proposed based on multiple features to obtain a fused degradation indicator to effectively describe bearings' degradation and avoid redundancy between the features. In [160], the fuzzy fusion method is used to obtain the fused degradation indicator from bearings vibration signals. Based on this indicator, the sub-modes of the degradation process are identified using the ensemble empirical mode decomposition. Finally, the extreme learning machine based model is used for predicting the bearings RUL.

In [161], a subset of multiple features is selected according to the Fisher linear discriminant coefficient. To take the advantage of the capability of capturing the system dynamic behavior quickly and accurately, the ANFIS is used as the individual feature models to constitute a distributed scheme to simplify the bearing RUL forecasting procedure, increase the generalization capabilities towards new data and reduce the prediction errors.

1.5 Concluding remarks

In general, the existing approaches for bearings RUL estimation are organized in two steps: 1) extract health indicators or indices for effectively describe the bearings degradation process and 2) establishing an accurate model for bearings RUL estimation.

Vibration signals are an essential source of bearing condition monitoring data due to a rather simple and cost-effective data collection process. Therefore, features extracted from vibration signals are always used for bearing condition monitoring. However, because of the complexity

of the bearings degradation evolution, common existing features have drawbacks such as insensitiveness to incipient defects and a highly fluctuating behavior along with the degradation evolution. In general, common features extracted from vibration signals are not monotonic and sensitive along the entire bearing lifecycle. Meanwhile, bearings are mechanical components and self-repair is improbable during the non-utilization periods. Moreover, the RUL along with the bearings' lifecycle is successively decreasing. Therefore, for monitoring the bearings' degradation, it is natural that a valuable characteristic of the new indicator is monotony through the entire bearing lifecycle. Unfortunately, the existing features and indicators are rarely monotonic during the degradation process. Besides, most of them are only sensitive to specific stages of the degradations.

Numerous approaches have been developed for bearings RUL estimation. Event data-based approaches do not require the condition monitoring data. They provide an estimation associated with the population information rather than the information associated with the individual components.

Several classes of condition data-based approaches were proposed:

1) physical Failure model approaches rely on the assumption that the physical failure model can describe well the degradation process. These approaches do not require a significant amount of condition data. Unfortunately, the degradation process of bearings is too complex and it is difficult to derive an accurate enough physical model.

2) The statistical data-driven approaches do not provide a physical failure model. They are poor performance drawback with high dimensional data and the requirement of presumed degradation indicator thresholds. Primarily, the assessment accuracy of statistical data-driven approaches strongly relies on the setting of presumed thresholds. The setting of presumed thresholds is not easy to be implemented because of the property of the high dispersion in the bearings RUL.

3) Artificial intelligence approaches do not need a formal model. They can learn from experience. Therefore, they can be used when it is difficult to derive a mathematical or statistical model. However, artificial intelligence models need a careful setting of the parameter, which has a strong influence on the accuracy of the results A significant amount of data is needed for training the model and obtaining the optimized values for the parameters of the models.

Among artificial intelligence frameworks, fuzzy logic triggered the attention of researchers due to its nature to process linguistic, ambiguous information and consider incomplete information.

Therefore, it might be suitable for modeling the ambiguous aspect of the bearing degradation evolution.

In the literature, due to the complexity of bearings degradation processes, different models have been used in various methods for estimating and forecasting bearings RUL. Thus, most of the models are composite and obtained using the bearings operating condition data. Although the results with the composite models are relatively accurate, they require a significant amount of data for deriving and identifying the models.

1.6 Research objectives

As in the aforementioned content, there are two steps in the approaches of bearings RUL estimation, i.e., indicators extraction and model identification.

The thesis that I developed in this PhD motivated by the current state of the art in bearings RUL estimation research.

1) Based on vibration signals, we aim to develop new degradation indicators which should be monotonic and sensitive through the entire bearing lifecycle. Additionally, by considering the historical information of the bearings degradation process, we worked out new indicators that will carry more useful information for bearings condition monitoring.

Multivariate analysis methods can serve to characterize diverse underlying correlation structures of Multivariate time series [162]. And, a suitable method is the so-called piecewise linear representation method, which directly provides features in the time-domain [163]. Based on the time-series of common features, we will use multivariate analysis methods and a piecewise linear representation method to build new degradation indicators for bearings RUL estimation in Chapter 2.

2) In the case of new type bearings, with historical data from a small available number of bearings, we aim to work out a method capable of identifying a simple structure model for estimating bearings RUL.

Some approaches do the prediction based on time series of degradation features or health indices. Then the RUL estimation is implemented using a predefined threshold, e.g., [152] [164]. But it is not easy to set an appropriate value of threshold because the dispersion of bearing life is tremendous. Hence, in this work, the model proposed will use the ratio associated with RUL as its output, which is similar to several models in the literature [140].

This work aims to find an effective way to identify a simple structure model with few parameters and relative accuracy for estimating bearings RUL under the condition of the small-size of the sample. The fuzzy theory has demonstrated its efficiency for the cases with incomplete information [156]. Moreover, fuzzy sets based modeling techniques have the capability of modeling complex nonlinear systems by decomposing the modeling problem into simpler linear sub-problems [165]. Among the classic fuzzy systems, the Takagi-Sugeno (T-S) Fuzzy Inference System (FIS) is computationally efficient and more suitable for mathematical analysis compared to the Mamdani FIS [166]. Thus, we use a multiple-input single-output (MISO) T-S FIS model for bearings RUL estimation. Base on small size training data sets, a new identification method of the MISO T-S FIS model for bearings RUL estimation is proposed in Chapter 3. Based on small size additional training data sets, a tuning approach of the MISO T-S FIS model for bearings RUL estimation is proposed in Chapter 4.

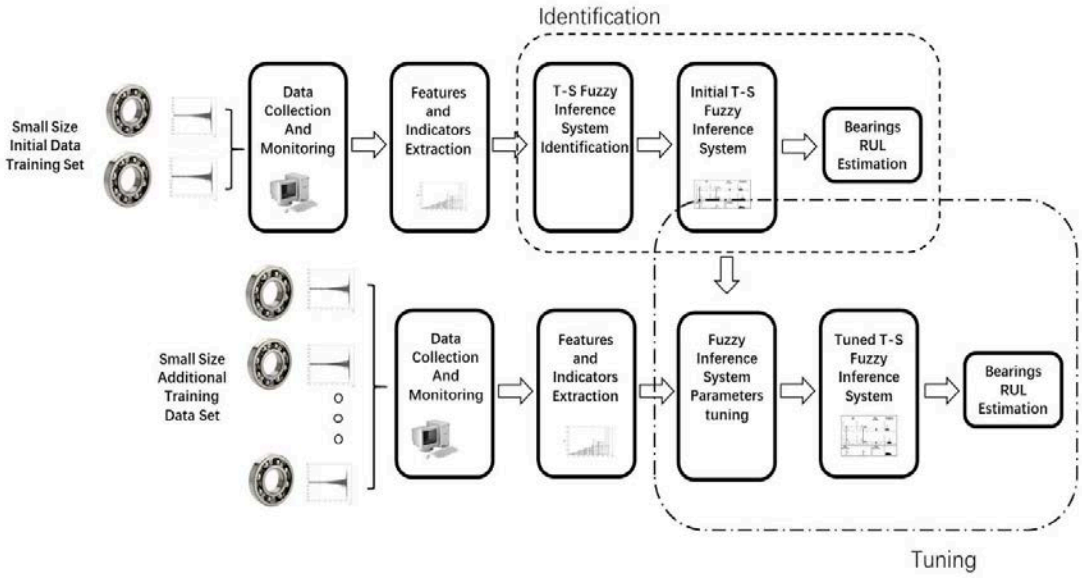


Figure 1.14 Overview of this work.

Chapter 2 - Synthesis of bearing degradation indicators

The accuracy of bearings degradation assessment depends largely on degradation indicators. Bearings cannot self-repair during periods of inactivity. In other words, the degradation evolution of bearings is cumulative and increasing. Therefore, a feature with an underlying monotonic trend, which increases continuously with the degradation cumulation, is helpful for monitoring bearings' degradation for CBM decision support. Such a feature is called a monotonic feature.

In this chapter, we present two methods that we elaborated to derive monotonic indicators for bearing degradation monitoring, based on common degradation features extracted from vibration signals. These methods are based on a piecewise linear representation approach and discarded space information to elaborate bearings degradation monitoring indicators, which are monotonic and sensitive during the degradation process of bearings. These results were communicated to three international conferences [167][168][169] and accepted for publication in the International Journal of Mechatronics and Automation.

2.1 Standard degradation feature

In general, the bearings vibration signal is effective in implementing data sources for revealing bearings' conditions. As we mentioned in Chapter 1, the probability density function of the vibration signal also changes along with the evolution of the bearing degradation process. Therefore, statistical time-domain features extracted through examining the probability density function of the signal can be used to assess the condition of bearings. Common statistical time-domain features are RMS, mean, STD, PMR, kurtosis and skewness.

Let us recall the common degradation features we mentioned earlier. The RMS, derived from the vibration signals, provides additional information on the energy quantity of the signal. It is an important feature used to estimate bearing degradation [19]. It increases typically gradually as fault developed. The definition of RMS is in (1.1). The mean quantifies the mean amplitude of signals, defined in formula (1.2). Note that, for raw vibration signals collected from bearings degradation, the mean remains close to zero but is not zero [20]. This is due to the inherently dynamic characteristic and structural defects of bearings, environmental noises, and irregular shifts of mechanical structures related to bearing degradation. Vibration signals in rolling bearings are non-stationary in nature[21][22]. The STD measures the dispersion of the signals around the mean value. Its definition is provided by the formula (1.3). The PMR quantifies the intensity of the pulses of the signal. Its definition is given in (1.4). A large PMR value corresponds to a vibration signal with a few large spiky peaks. Similarly, a low PMR value is relevant to a flattish vibration signal [23]. The mean and standard deviation values of the distribution are the first two parameters of the sample distributions. The first third parameter of the distribution is skewness and the first fourth being kurtosis. The kurtosis reveals the degree of deviation of the signal distribution from the average and the skewness indicates the signal distribution asymmetry, their definitions are respectively in (1.5) and (1.6).

Bearings are mechanical components that cannot self-repair during periods of inactivity. In other words, the degradation evolution of bearings is cumulative and increasing. Therefore, a feature with an underlying monotonic trend, which increases continuously with the degradation cumulation, is helpful for monitoring bearings degradation for CBM decision support. Such a feature is called a monotonic feature. Unfortunately, common degradation features extracted from bearings vibration signals are not monotonic during the degradation process. Each common degradation feature is sensitive to specific degradation stages. Numerical examples are given in chapter 5.

To overcome this problem and build a degradation indicator monotonic with the evolution of the degradation, we propose to merge six common statistical time-domain features as the standard degradation feature F_k for describing the degradation process of bearings, as in (2.1).

$$F_k = [RMS_k, mean_k, STD_k, PMR_k, kurtosis_k, skewness_k]^T, \quad k = 1,2,\dots,K \tag{2.1}$$

Based on this statistical time-domain feature, we elaborate on two bearing degradation indicators by using a piecewise linear representation approach and discarded space information.

2.2 Piecewise linear representation

A time series is a collection of observations made chronologically. According to the number of variables, time series can be categorized into two types: univariate and multivariate time series. During the bearing degradation process monitoring, the standard degradation features F_k, collected chronologically, form a multivariate time series.

Recently, there has been a high interest in the field of data mining, especially for the time series. However, elaborating characteristic features to represent the time series data is the staple issue. In [170], the representation of time series is implemented by transforming the time series to another domain, used for indexing time series. Based on a local view of the database, a novel framework is introduced using ensembles of two or more representations for more efficient indexing. It employs different approaches to represent different parts of the database. The local view of the database is that parts execute a high fidelity representation using wavelets or Fourier transforms according to effectiveness. In [171], discrete Fourier transforms are used to carry on the representation of time series for similarity measurement. In [172], orthonormal and bi-orthonormal wavelets are successfully used to represent the time series for similarity search. An empirical study of similarity search discussed the performance of a large number of wavelet methods. In [173], a novel wavelet-based tree structure is proposed to obtain a multi-level representation of time series. In [174], a new method is proposed using both a discrete Fourier transform and a wavelet transform to conduct time series representation.

The above methods represent a time-series in the transformation domain. The basic idea is to approximate the original time series with a few transform coefficients to map them into low-dimensional space.

Other methods provide an approximation of the time series while representing time series in the time-domain. A popular method for time series representation is the piecewise linear representation method. The piecewise linear approximation method refers to the approximation of a time series with straight lines (or constants), where the number of straight lines (or

constants) is much less than the size of the original time series. The representation based on piecewise linear approximation is a simple form comparing with the original time series and makes the storage, transmission and computation of the data more efficient. In the field of time series data mining, the piecewise linear approximation has been adapted to various applications. The simplest method is sampling described in [175]. By adopting an appropriate sampling interval, the time series can be represented along with dimensionality reduction. However, the sampling method has the drawback of distorting the shape of the time-series where the sampling interval is too big, according to the Nyquist sampling theorem. In [176], an enhanced method is proposed. First, each time series is partitioned into several segments of equal length. Next, the mean is calculated from each segment, which is treated as a feature for each segment. The time series is represented using the segmented means. In [177], a similar transformation is proposed to approximate the time series data by segmenting the sequences into equal-length sections and recording the mean value of these sections. These mean values can then be indexed efficiently in a lower dimensionality space to support fast and exact similarly search.

For the features of segments, in [178], so-called segmented sum of variation is employed as the feature of each segment. For the length of segments, in [179], an extended version is proposed called an adaptive piecewise constant approximation, in which the length of each segment is not fixed, but adaptive to the shape of the series.

Besides, in [180], based on segmenting sequences into homogenous segments, a compact representation derived from each segment as a mathematical function is introduced. Indexing time series is according to the properties of behavior captured from the functions. In [181], an extended representation of the time-series is proposed to fast, accurate classification and clustering. The representation consists of piecewise linear segments and weight vectors. The piecewise linear segments are used to represent the shape, and the weight vectors reveal the correlations of each linear segment.

Basically, the piecewise linear representation method is to segment time series into several segments then use straight lines (or constants) as the characteristic features of the segments in order to approximately represent the entire time series. Obtaining the piecewise linear representation of time-series relies on two tasks: the time-series segmentation, and subsequent representation. Therefore, the implementation of a piecewise linear representation method relies on segmentation algorithms for time series and characteristic measurements of segments.

2.3 Segmentation algorithm of time series

The multivariate time series F consists of the standard degradation features collected chronologically. It is evolving through the bearing degradation process.

There are several main origins of bearing failure, i.e., excessive loading, overheating, false brinelling, true brinelling, normal fatigue failure, reverse loading, contamination, lubricant failure, corrosion, misalignment, loose fits, tight fits [10]. Each of the causes can result in the formation of particular damage that is specific to the corresponding degradation mechanism and can leave its own particular impact on the bearing. In other words, during the degradation process, the degradation occurs in each component of bearings, and several kinds of symptoms might meet. Hence, the bearing degradation process is complex and dynamic.

The bearing degradation process is usually characterized by several degradation stages. These stages embody in the multivariate time series F as the homogeneous segments. It exists a qualitatively significant difference in the behavior of the degradation process associated with different degradation stages. This evolution from one degradation stage to another is reflected in the multivariate time series F by inhomogeneous behaviors such as changes of the numerical scale, distribution and correlation structure.

The segmentation algorithms of time series aim to divide the time series into several homogeneous segments. There are a variety of segmentation algorithms of time series existed that are used in a wide variety of applications. These algorithms appear in the literature under a variety of names, even if there are only subtle differences between them. In [163], an excellent overview of the segmentation algorithms of time series is given. In general, authors in [163] define three categories for segmentation algorithms of time series which can involve most segmentation algorithms of time series existed in the literature. The three categories are Top-Down, Bottom-Up, and Sliding Windows. Figure 2.1 to Figure 2.3 are the general descriptions of them.

The general framework of Top-Down algorithms is given in Figure 2.1. They rely on recursively dividing time series into subsequences. It stops when the predefined stopping criteria is satisfied. In [180], an algorithm belonging to the Top-Down category is proposed to segment time series into meaningful segments. Then the approximation of the time series is built based on the segments. In [182], an improved version of the algorithm proposed in [180] is introduced. Comparing with the original version, the improved version has less algorithmic complexity by defining more parameters and filtering out the uninteresting segments.

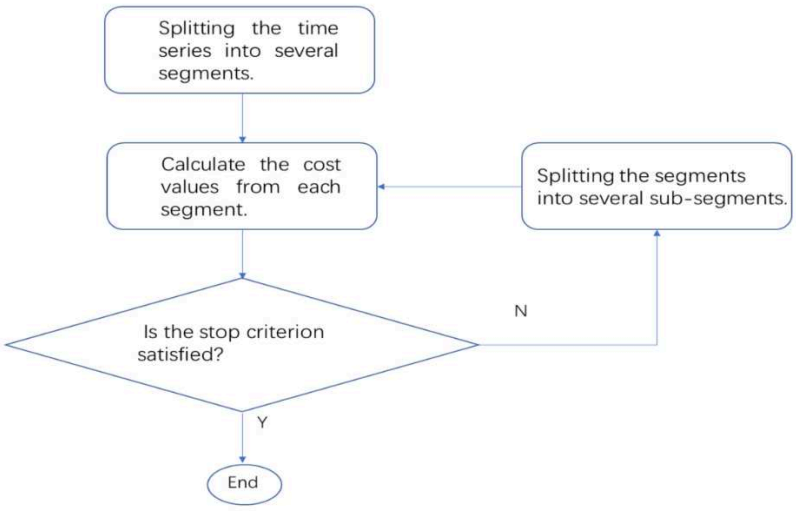


Figure 2.1 General description of Top-Down algorithm.

The Bottom-Up algorithms, illustrated in Figure 2.2, aim to divide time series into an ideal number of segments at the beginning, then iteratively merge the adjacent segments which are characterized by the smallest cost values in pair. In [183], an algorithm belonging to the Bottom-Up category is proposed to identify the events that are the segments in the data time series observed from the regular monitoring during intensive care through several channels. In [184], the algorithm of the Bottom-Up category is used with prior knowledge for matching patterns of time-series datasets.

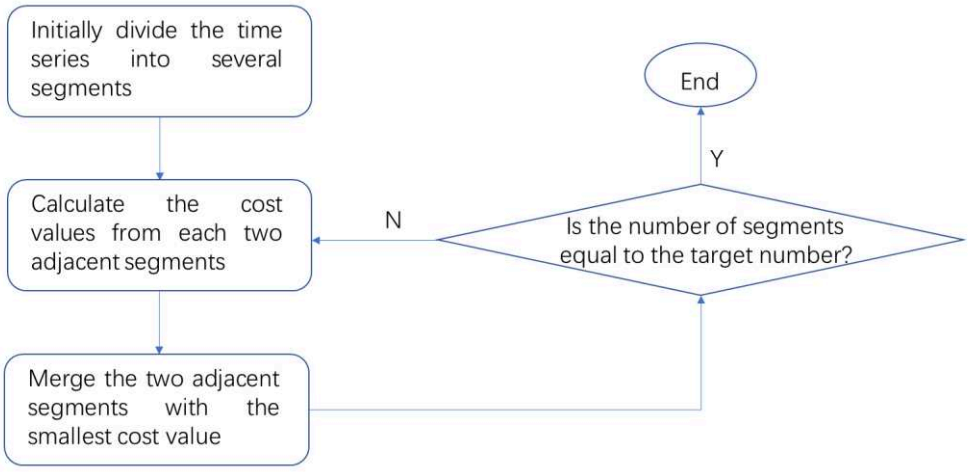


Figure 2.2 General description of the Bottom-Up algorithm.

The Sliding Windows algorithms, illustrated in Figure 2.3, rely on extending the segments until their cost value exceeds the predefined threshold. In [185], an algorithm belonging to Sliding Windows category is proposed to approximate (i.e., compress) digital electrocardiograms in

real-time. In [186], an algorithm of Sliding Windows category is used with a combination of several symmetrical distance functions to obtain the segments (i.e., sub-patterns) of the time series of electrocardiograms.

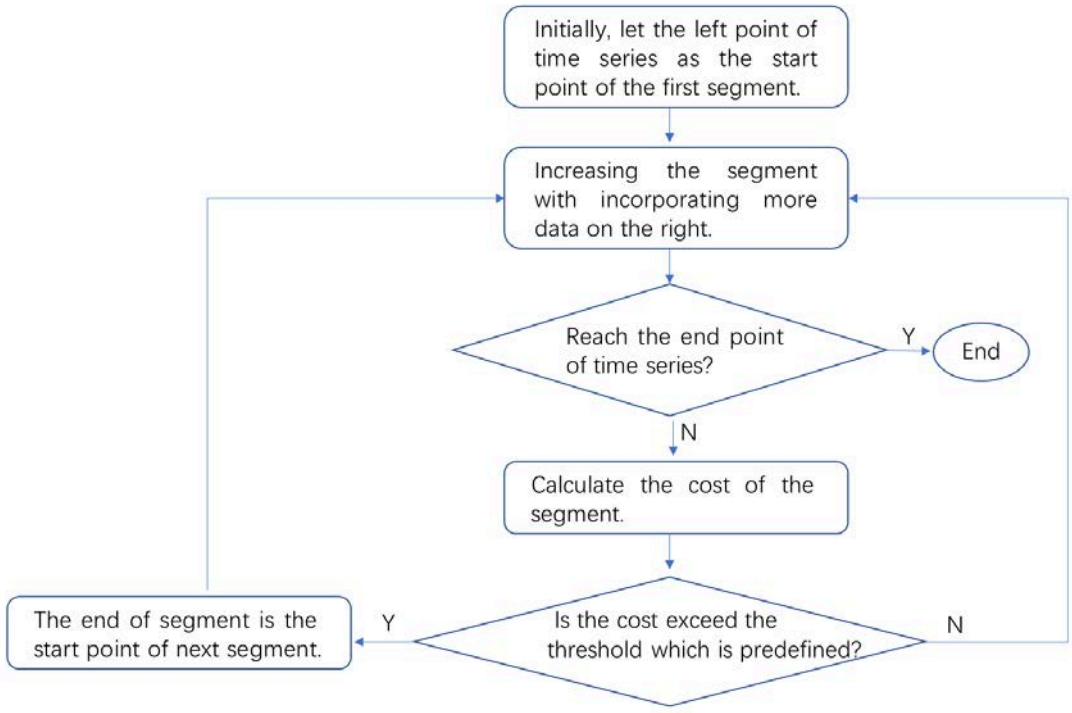


Figure 2.3 General description of Sliding Windows algorithm.

According to the application situation, the segmentation algorithms of time series can be categorized into two groups: batch algorithms and online algorithms [187]. On the one hand, for batch algorithms, the entire time series that need to be segmented must be available. On the other hand, online algorithms can consider new data in real-time, changing the boundaries of segments adaptively. Only the Sliding Windows algorithm can be used for online applications. A comparative presentation of the characteristics of these three categories of segmentation is provided in [163]. For the Bottom-Up algorithms and the non-recursive Top-Down algorithms, the maximum cost for each segment, and the number of final segments can be defined before execution. On the contrary, for the Sliding Windows algorithms, only the maximum cost for each segment can be defined before execution. The highest complexity characterizes the Top-Down algorithms among these classes of segmentation algorithms.

2.4 Principal component analysis for multivariate analyses

The segmentation algorithms summarized in the previous section and illustrated in Figure 2.1 to Figure 2.3 rely on cost values that need to be specified. The cost values are used to evaluate the changing points between adjacent homogeneous segments of time series. Additionally, the piecewise linear representation incorporates two tasks: time series segmentation and subsequence representation. The means for subsequence representation also needs to be specified.

One technique for finding the changing points and the subsequence representation is measuring the dissimilarity of multivariate time series. The dissimilarity of multivariate time series can be carried out in two different ways. One is to directly compare the variables of the multivariate time series to obtain the weight of them. This approach is intuitive and straightforward, but it is not as effective as we need. Since the multivariate time series are characterized both by the numerical value of its elements and by the relation between them. This relation corresponds to the correlation between the variables and it is considered as the hidden process, which carries the real description of a complex system.

In multivariate time series, multivariate analysis methods can serve to uncover the overlapping information among the correlated variables and characterize these diverse underlying correlation structures [162]. Principal component analysis (PCA) method is the most commonly used multivariate analysis method, which identifies underlying principal components among a set of variables and helps to understand the relations among variables. Through the PCA method, the data in the original space can be projected by orthogonal transformation into a new space with the same variable dimension. The new space is the full projected space. The data in the projected space is the projected space data called principal components.

The PCA method process aims to achieve a linear combination of the variables with maximized variance. The maximal variance relates to an essential aim that is to maximally spread out the observations in order to decouple correlation among the variables.

The PCA method process can be briefly described as follows: along the dimension with maximal variance, a linear combination generates the first principal component. The next principal component is generated by a linear combination along the dimension with maximal variance, which is orthogonal to the previous principal component. Then, to repeat the step, the following principal component can be generated.

In this work, the object being processed by the PCA method is the standard degradation feature F. So, the original space data is the time series F. Let φ denotes the projected space data i.e.,

principal components. The covariance matrix of F is denoted by Θ, which is subject to (2.2). The covariance matrix of φ is denoted by Ξ, which is subject to (2.3). $\aleph$ is defined as an orthogonal matrix, which is subject to (2.4). The expression (2.5) is obtained by substituting (2.4) into (2.3). Then, (2.6) is obtained by substituting (2.2) into (2.5). Let F^* denote the reconstructed original space data, which obtained using expression (2.7).

$$\Theta = \frac{1}{K} F F^T \tag{2.2}$$

$$\Xi = \frac{1}{K} \varphi \varphi^T \tag{2.3}$$

$$\varphi = \aleph F \tag{2.4}$$

$$\Xi = \aleph(\frac{1}{K} F F^T)\aleph^T \tag{2.5}$$

$$\Xi = \aleph \Theta \aleph^T \tag{2.6}$$

$$F^* = \aleph^{-1} \varphi \tag{2.7}$$

Since the variables in φ are uncorrelated, it is certain that Ξ is a diagonal matrix, which is described in (2.8). And $\aleph$ is an orthogonal matrix.

$$\Xi = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_I \end{bmatrix} \tag{2.8}$$

where i=1, 2…, I is the variable label of F. According to the basic theory of linear algebra, ψ_1, ψ_2, …, ψ_I are the variances corresponding to the variables of φ and $\aleph$ is the matrix whose columns are the corresponding eigenvectors of the matrix Θ. Thus, the PCA method can be implemented by (2.4).

A critical application of the PCA method and its derivatives is the batch process characterization in production [188]. Several derivatives of the PCA method are proposed in the literature to uncover correlations to improve the characterization of multivariate time series of batches, e.g., [189] [190][191]. All such literature proves that the PCA method is capable of identifying and characterizing the segments in multivariate time series.

2.5 Synthesis of bearing degradation indicators based on segmented discarded projected space information

Considering the limitations of the standard features mentioned in the previous section for monitoring the bearings degradation, a valuable characteristic for the new indicator is monotonicity through the entire bearing lifecycle. This requirement is natural, knowing that bearings are mechanical components and self-repair is improbable during the non-utilization periods.

As bearings degradation is a complex process, we think that enhancing the indicator by considering the information on the evolution of the bearing's degradation process will increase its effectiveness.

2.5.1 **Discarded projected space information**

The bearings degradation influences the underlying correlation structure of the selected common features. Therefore, we suggest using the underlying correlation structures of the selected common features to track the bearings' degradation. Multivariate analysis methods can serve to characterize these diverse underlying correlation structures [162].

PCA method is the most commonly used multivariate analysis method, which identifies underlying principal components among a set of variables and helps to highlight the relations among variables.

Among all the obtained principal components through the transformation of PCA method, the retained principal components embody most of the variance of the original data [162]. Except retained principal components among all the obtained principal components, they are the discarded principal components. All the obtained principal components constitute the full projected space. The retained principal components constitute the reduced projected space. The difference between the full projected space and reduced projected space is called discarded projected space in this work.

In this work, we employed the cumulative percent variance to fix the number of principal components retained to describe the data effectively [192]. Commonly, retained principal components with the higher variances are used to describe the original data and reduce the space dimension for the data analysis. However, the discarded principal components with smaller variances may also carry useful information to some extent [162]. To that extent, we propose using the discarded Hotelling T square to measure the discarded projected space information, which is associated with the difference between the full projected space and the reduced projected space to track the bearings' degradation.

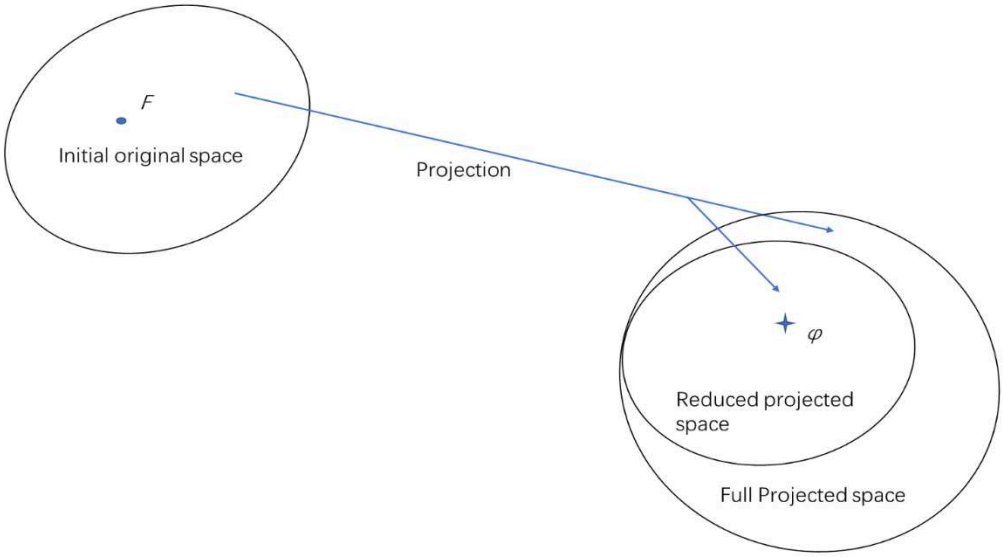


Figure 2.4 The relationships among different projected spaces through the PCA method.

Figure 2.4 depicts the relationships among the full original space, the full projected space and the reduced projected space through the PCA method, where F is the initial original space data, φ is the projected space data.

2.5.2 **Segmented discarded Hotelling t square**

The Hotelling T square is the measure of the multivariate distance between each value and the mean of the data set. The squared Mahalanobis distance is employed to estimate the Hotelling T square [193] In this work, we propose the segmented discarded Hotelling T square to survey the multivariate distance in the projection space to track the bearings degradation.

Comparing with Euclidean distance, Mahalanobis distance measurement considers the variance and covariance of the variables when calculating the distances, which has the advantage that the principal components are equally weighted during the calculation. One can capture all changes that happen in the components even with small variances [194]

The Hotelling T square is defined in (2.9) as it follows

$$T^2_{m,k_m} = (\varphi_{m,k_m} - \vartheta_m)\phi_m^{-1}(\varphi_{m,k_m} - \vartheta_m)^T \tag{2.9}$$

with

$$\varphi_{m,k_m} = \left[\varphi^1_{m,k_m}, \varphi^2_{m,k_m}, \cdots, \varphi^\beta_{m,k_m}\right] \tag{2.10}$$

$$\vartheta_m = \left[\vartheta^1_m, \vartheta^2_m, \cdots, \vartheta^\beta_m\right] \tag{2.11}$$

where $m=1,2,\ldots, M$ is the segment label, $k_m=1, 2,\ldots,K_m$ is the values label of the m^{th} segment, $i=1,2,\ldots,I$ is the label of variable, ϕ_m is the covariance matrix of the first β principal components in the m^{th} segment, φ^{i}_{m,k_m} is the $k_m{}^{\text{th}}$ value of the i^{th} principal component in the m^{th} segment, ϑ^{i}_{m} is the mean of the i^{th} principal component in the m^{th} segment.

When β is equal to the variable dimension I of the original data, T^2_{m,k_m} in formula (2.9) is the full projected space Hotelling T square TF_{m,k_m}. When β is equal to p, T^2_{m,k_m} in formula (2.9) is the reduced projected space Hotelling T square TR_{m,k_m}. p indicates how many principal components are treated as the retained principal components.

The segmented discarded Hotelling T square Λ^2_m, proposed in this work, is the Discarded Hotelling T square value of a time series segment, defined in equation (2.12). It is a measurement of the discarded projected space information associated with segments of multivariate time series.

$$\Lambda^2_m = \frac{1}{K_m}\sum_{k_m=1}^{K_m} TF_{m,k_m} - TR_{m,k_m} \tag{2.12}$$

2.5.3 **Indicators synthesis**

Indicator SDHT²

Bearings degradation is a complex process. Therefore, we believe that it is valuable for a bearing degradation indicator to carry information on the historical degradation process. We use discarded information through the PCA method to implement this idea to track the bearings' degradation.

Suppose that a multivariate time series F was continuously collected along with the operation time of a bearing. The beginning sample interval of the bearing using life is k_0, and its final sample interval of the multivariate time series F is k_{end}. Let k_i corresponding to an arbitrary sample interval between k_0 and k_{end}. Considering the calculation of Λ^2 should involve an amount of data, the indicator estimation starts at a sample interval k_{start}, the bearing during the life duration before the sample interval k_{start} is considered with the perfect condition. The complete run to failure bearings could be used as the training set to predefine the value of k_{start}. The predefinition of k_{start} is set as corresponding to 5% of the average of all training bearings lifetime.

Note that: the value of Λ^2 derived from the multivariate time series F from k_0 to k_i are respectively to be the value of the indicator SDHT^2 corresponding to the time point k_i.

The SDHT^2 values corresponding to each time point after k_{start} of the entire time series F is denoted by SDHT$^2_{series}$.

Algorithm 2.1 SDHT$^2_{series}$ extraction

Initial step:

$k_i = k_{start}$ ($k_0 < k_{start} < k_{end}$, k_{start} is a constant);

q=1;

While $k_i \leq k_{end,}$

Call formula (2.12) to calculate the Λ^2_m value based

on the segment of time series from k_0 to k_i;

SDHT$^2_{series}$ (1, q) =Λ^2;

$k_i = k_i$+1;

$q = q$+1;

end

Indicator VSDHT^2

The indicator SDHT^2 presented in the previous section carries the information of the evolution of the bearing's degradation process. It treats the historical process as a single block without considering the historical process has multiple stages.

If one considers that the historical process has multiple stages, we need to distinguish each stage of the historical process then use Λ^2 to respectively characterize each stage of the historical process. Hence, we propose using a segmentation algorithm of time series to divide the historical data of the multivariate time series F into homogeneous segments corresponding to the stages of the historical process to implement the piecewise linear representation.

As mentioned in the previous section, several algorithms are proposed in the literature for segmenting time series. Each group of segmentation algorithms has its specific advantages and disadvantages. The Sliding Windows algorithms are not able to divide time-series data into a predefined number of segments. Top-Down algorithms are more complex than the Bottom-Up algorithms [195].

To build indicators from multivariate time series F, we segmented the time series into homogeneous data segments related to different degradation stages. As a requirement, the indicators must have the same dimension.

A Bottom-Up algorithm seems to be more suitable for our investigations because of lower complexity and the possibility to fix the desired final number of segments.

Since the capability of the PCA method described in the previous section, segmented discarded Hotelling T square can be used to segment the time series data into homogeneous sections by tracking the changing of the underlying correlation structure among the multivariate time series.

a) The Extraction Algorithms

The $VSDHT^2$ health indicator is defined in equation (2.13). The pseudo-code of the Bottom-Up algorithm for extracting the indicator is presented hereafter.

$$VSDHT^2 = \left[\Lambda_1^2, \Lambda_2^2, \ldots, \Lambda_Q^2\right]^T \tag{2.13}$$

Algorithm 2.2 Indicator $VSDHT^2$ extraction

1) Let l be the length of a multivariate time series F from t_0 to t_i.

2) Fix the final number of segments Q and the parameter z used to define the initial segments of the time series.

3) Initializing: Separate the time series F into z segments of equal length round(l/z), except for the last segment's length that is rem(l/round(l/z)).

where $l>z>Q$

4) While $z>Q$,

Combine every two adjacent segments and calculate the value of segmented discarded Hotelling T square of each combination as the cost values;

Merge the two adjacent segments with the smallest value of segmented discarded Hotelling T square;

$z=z-1$;

end (while).

5) Calculate the value of segmented discarded Hotelling T square of each segment as the characteristic values. The vector with Q entries is the indicator $VSDHT^2$.

Note that:

(1) the value of indicator $VSDHT^2$ derived from the multivariate time series F from t_0 to t_i is corresponding to the time point t_i.

(2) when calculating the segmented discarded Hotelling T square as the cost value, the labels m and k_m in formulas (2.12) are associated with the segments consisting of adjacent segments. When calculating them as the characteristic value, the labels m and k_m in formulas (2.12) are associated with final segments of the time series.

The value of $VSDHT^2$ corresponding to each time point of the entire time series of a bearing is denoted by $VSDHT^2_{series}$. The pseudo-code for the $VSDHT^2_{series}$ extractions are in Algorithm 2.3.

Algorithm 2.3 $VSDHT^2_{series}$ extraction

Initial step:

$k_i = k_{start}$ ($k_0 < k_{start} < k_{end}$, k_{start} is a constant);

q=1;

While $k_i \leq k_{end,}$

Call Algorithm 2.2 to calculate the VSDHT^2 value based on the segment of the time-series from k_0 to k_i;

VSDHT$^2_{series}(1, q)$= VSDHT^2;

$k_i = k_i+1$;

$q = q+1$;

end

Based on training bearings complete run to failure, the setting of the parameters z , Q of the Algorithm 2.2 as well as parameter p is discussed in the next two parts of this section.

b) Parameters of The Extraction Algorithm

The parameters of the extraction algorithm are the parameters z and Q of the Algorithm 2.2. Spearman rank correlation coefficient (SRCC) is a nonparametric technique for evaluating the degree of linear association or correlation between two independent variables [196]. We respectively calculate the SRCC between the time order series and the VSDHT^2 data series of each variable dimension from every training sample. We use the average standard deviation of SRCC (ASDS) defined in formula (2.14) to compare the VSDHT$^2_{series}$ performance for different values of parameters z and Q.

$$ASDS = 1/S \sum_{s=1}^{S} \sqrt{1/(Q-1) \sum_{q=1}^{Q} (SRCC_q{}^s - (1/Q \sum_{q=1}^{Q} SRCC_q{}^s))^2} \tag{2.14}$$

where $s=1, 2, ..., S$ is the label of the training bearings, $q=1, 2, ..., Q$ is the label of variables in the VSDHT^2 indicator.

The ASDS measures the divergence between different variables dimensions of VSDHT$^2_{series}$ data series associated with the amount of information carried by the indicator. The bigger value of ASDS, the more effective VSDHT^2. By comparing the ASDS values based on different z and Q values, we choose the VSDHT$^2_{series}$ that carries more information. Thus, we choose the values of z and Q that correspond to the largest value of ASDS.

c) Cumulative percent variance (CPV)

In order to calculate the values of Λ_m^2, we need to set the number p of retained principal components.

$$CPV = \left(\frac{\sum_{i=1}^{p} \psi_i}{\sum_{i=1}^{I} \psi_i} \right) \times 100\% \tag{2.15}$$

where ψ_i denotes the variance of i^{th} principal component.

Ranking the principal components φ with corresponding variance ψ in descending order, CPV is the measurement of the percent variance extracted by the first p principal components as defined in equation (2.15) [192].

The selected number p of retained principal components is set such that the CPV value is equal at least to 95% for every training bearing.

2.6 Synthesis of bearing degradation indicators based on segmented discarded original space information

2.6.1 Discarded original space information

The new indicators presented in the previous section are based on discarded projected space information. As we know, through the PCA method, the reduced reconstructed original space could be reconstructed by using retained principal components. The discarded original space information, which is associated with the difference between the initial original space and the reduced reconstructed original space, is used in this section to track the bearings' degradation.

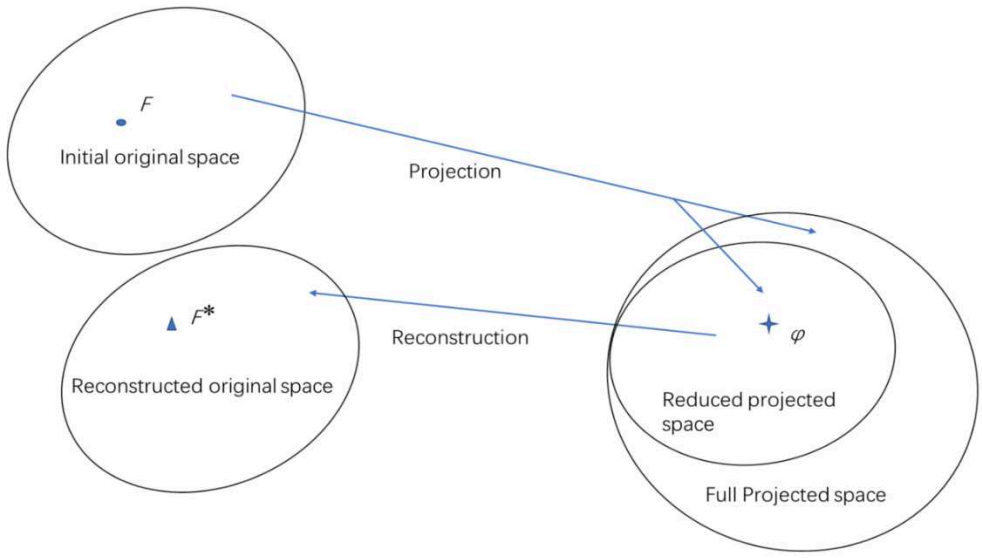


Figure 2.5 The relationships among different spaces through the PCA method.

Figure 2.5 depicts the relationships among the initial original space and the reduced original space through the PCA method, where F^* is the reconstructed original space data.

2.6.2 Segmented projected error

The segmented projected error Δ^2_m proposed in this work is the average squared distance between the initial original space and the original reconstructed original space, through the PCA method, as defined in equation (2.16). It is a measurement of the discarded original space information associated with segments of multivariate time series.

$$\Delta^2{}_m = \frac{1}{K_m}\sum_{k_m=1}^{K_m}\sum_{i=1}^{I}\left\| F_{m,k_m[i]} - F^{p*}_{m,k_m[i]} \right\|^2 \quad (2.16)$$

where F^{p*} is the reduced reconstructed original data by p retained principal components, as illustrated in Figure 2.5.

First, we rank the principal components φ with corresponding variance ψ in descending order and then let φ^p denote the matrix consisting of the first p rows of φ. The corresponding columns of matrix $\aleph$ constitute the matrix $\aleph^p$. According to formula (2.7), F^{p*} can be obtained by formula (2.17).

$$F^{p*} = (\aleph^p)^{-1}\varphi^p \quad (2.17)$$

Using the segmented projected error $\Delta^2{}_m$ to replace Λ_m^2 in the process of the previous section, two more new indicators are presented in this section.

2.6.3 **Indicators synthesis**

Indicator SPE

The pseudo-code of the SPE_{series} extraction is Algorithm 2.4

Algorithm 2.4 SPE_{series} extraction

Initial step:

$k_i = k_{start}$ ($k_0 < k_{start} < k_{end}$, k_{start} is a constant);

q=1;

While $k_i \leq k_{end,}$

Call formula (2.16) to calculate the Δ^2 value based on the segment of time series F from k_0 to k_i;

$SPE_{series}(1, q) = \Delta^2$;

$k_i = k_i + 1$;

$q = q + 1$;

end

Indicator VSPE

The indicator SPE presented in the previous section carries the information of the evolution of the bearing's degradation process. It treats the historical process as a single block without considering the historical process has multiple stages.

To consider that the historical process has multiple stages, $\Delta^2{}_m$ is used to distinguish each stage of the historical process and characterize each stage of the historical process. The Bottom-Up algorithm is used to divide the historical data of the multivariate time series F into homogeneous

segments corresponding to the stages of the historical process to implement the piecewise linear representation.

The VSPE health indicator is defined in equation (2.18). The pseudo-code of the Bottom-Up algorithm for extracting the indicator VSPE is presented hereafter.

$$VSPE = [\Delta^2{}_1, \Delta^2{}_2, \ldots, \Delta^2{}_B]^T \tag{2.18}$$

Algorithm 2.5 Indicator VSPE extraction

1) Let l be the length of a multivariate time series F from t_0 to t_i.

2) Fix the final number of segments B and the parameter u used to define the initial segments of the time series.

3) Initial: Separate the time series F into u segments of equal length round(l/u), except for the last segment's length that is rem(l/round(l/u)).

where $l>u>B$

4) While $u>B$,

Combine every two adjacent segments and calculate the value of segmented discarded Hotelling T square of each combination as the cost values;

Merge the two adjacent segments with the smallest value of segmented discarded Hotelling T square;

$u=u-1$;

end (while).

5) Calculate the value of the segmented projected error of each segment as the characteristic values.

The vector with B entries is the indicator VSPE.

Note that:

(1) the value of indicator VSPE derived from the multivariate time series F from t_0 to t_i is corresponding to the time point t_i.

(2) when calculating the segmented projected error as the cost value, the labels m and k_m in formula (2.16) are associated with the segments consisting of adjacent segments. When calculating them as the characteristic value, the labels m and k_m in formula (2.16) are associated with final segments of the time series.

The values of VSPE values corresponding to each time point of the entire time series of a bearing is denoted by $VSPE_{series}$. The pseudo-code for the $VSPE_{series}$ extraction is in Algorithm 2.6.

Algorithm 2.6 $VSPE_{series}$ extraction

Initial step:

$t_i = t_{start}$ ($t_0 < t_{start} < t_{end}$, t_{start} is a constant);

q=1;

While $t_i \leq t_{end}$,

Call Algorithm 2.5 to calculate the VSPE value based on the segment of time series from t_0 to t_i;

$VSPE_{series}(1, q)$= VSPE;

$t_i = t_i+1$;

$q = q + 1$;

end

Base on the training bearings, which are complete run to failure, the setting of the parameters u and B of the Algorithm 2.5 can be selected by formula (2.19), which is similar to the formula (2.14).

$$ASDS = 1/S \sum_{s=1}^{S} \sqrt{1/(B-1) \sum_{b=1}^{B} (SRCC_b{}^s - (1/B \sum_{b=1}^{B} SRCC_b{}^s))^2} \quad (2.19)$$

where $b=1, 2, \ldots, B$ is the label of variables in the VSPE indicator.

The values of SRCC are calculated between the time order series and the $VSPE_{series}$ of each variable dimension from every training sample. The ASDS measures the divergence between different variables dimensions of $VSPE_{series}$ data series associated to the amount of information carried by the indicator. The bigger value of ASDS, the more effective VSPE. By comparing the ASDS values based on different u and B values, we choose the $VSPE_{series}$ that carries more information. Thus, we choose the values of u and B that correspond to the largest value of ASDS. In order to calculate the values of Δ^2, we need to set the number p of retained principal components. Ranking the principal components φ with corresponding variance ψ in descending order, the selected number p of retained principal components is set such that the CPV value is equal at least to 95% for every training bearing.

2.7 Summary

In this work, we proposed four new bearing degradation monitoring indicators $SDHT^2$, SPE, $VSDHT^2$, and VSPE. They all can uncover the underlying correlation structure of the multivariate time series of standard degradation features by using discarded space information to track the degradation evolution of bearings. Moreover, they all highlight the historical degradation information rather than only the current condition. Especially, the indicators $VSDHT^2$ and VSPE, using piecewise linear representation, highlight the historical degradation information, considering the historical process has multiple degradation stages. For the indicators $VSDHT^2$ and VSPE, they represent different aspects of the bearings degradation process due to their different underlying measurements.

Identical bearings which are complete run to failure are used as a training set to obtain the parameters to the extraction process of these indicator proposed in this chapter. Then, according to these parameters, the indicators could be real-time extracted along with the lifetime of test bearings. The indicators extracted from the over the complete run-to-failure cycle of training

bearings will be used to identify models for bearings RUL estimation. The real-time indicators extracted along with the lifetime of test bearings will be the input of the models to estimate the test bearings RUL.

The model identification and tuning approaches will be proposed in chapter 3 and 4. For the new indicators proposed in this chapter, all the properties of monotonic, continuous sensitivity, and effectiveness for bearings RUL estimation will be analyzed in chapter 5.

Chapter 3 - Fuzzy model identification based on small-size training datasets mixture distribution analysis for bearings remaining useful life estimation

The research work presented in this chapter proposes a data-driven modeling method for bearings remaining useful life estimation based on Takagi-Sugeno (T-S) fuzzy inference system (FIS). This method allows identifying the parameters of a classic T-S FIS using features extracted from periodic observation of the vibration signals collected from a small number of bearings over the entire run to failure period. The number of rules and the input parameters of each rule are identified using the subtractive clustering method. Furthermore, we propose to use the maximum likelihood method of mixture distribution analysis to calculate the parameters of clusters on the time axis and the probability corresponding to each degradation stage. Based on this result, we identified the output parameters of each rule using a weighted least square estimation. This result was valorized through a journal paper submitted for publication to Mechanical Systems and Signal Processing. The current status in the submission system is under review.

3.1 Problem statement

The research work introduced in this chapter aims to elaborate an effective method to identify a classic simple structure T-S FIS with a small number of parameters as the model for estimating bearings RUL using features extracted from the vibration signals observed over a complete run-to-failure cycle from a small number of identical bearings operating under identical load conditions. We assume that there is no available information on the failure modes nor a fixed failure threshold.

Figure 3.1 illustrates the definitions of the PUL and the RUL associated with the bearings exploitations.

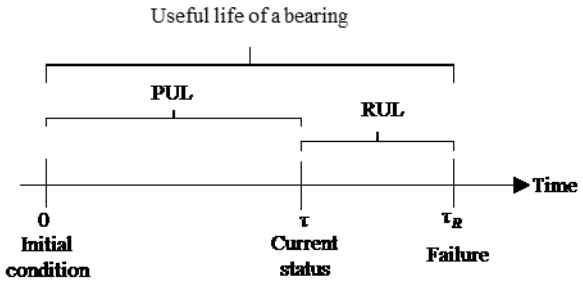


Figure 3.1 Bearing useful life.

Let ρ define the consumed useful life ratio of a bearing in equation (3.1).

$$\rho = \frac{PUL}{PUL + RUL} = \frac{\tau}{\tau_R} \tag{3.1}$$

where τ denotes the PUL which is known and corresponds to the period that the bearing was used until the current observation, τ_R denotes the overall lifetime, which is the time over a complete run-to-failure cycle of the bearing.

Using relation (3.1), one can calculate the RUL define by $\Delta\tau_R = \tau_R - \tau$ based on the PUL and the consumed useful life ratio using (3.1), as it follows:

$$\Delta\tau_R = \left(\frac{1}{\rho} - 1\right)\tau \tag{3.2}$$

Let V_k=[$v_{k,1}$, ..., $v_{k,I}$]^T denotes the input vector of the model, which is constituted by the features or indicators extracted from the vibration signals of k^{th} sample interval during bearings monitoring, where k=1,2,...,K, and i=1,2,...,I is the variable label. The corresponding output of the model is the estimated consumed useful life ratio denoted by $\hat{\rho}_k$. The model is summarized in Figure 3.2.

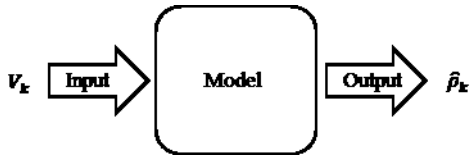


Figure 3.2 The model for RUL estimation.

In this work, the input vector V_k is constituted by the features or indicators extracted from the vibration signals of k^{th} sample interval during bearings monitoring. Takagi-Sugeno fuzzy inference system for bearings RUL estimation

3.1.1 The input and output variables model

We use FIS to establish the model by identifying the sample data with the corresponding number of rules. Each data cluster corresponds to a degradation specific stage in the space of state features of the bearings. For that purpose:

1) Let's consider the degradation as a fuzzy set S defined on a universe of discourse of lifespan, and let's consider a finite number of fuzzy subsets s_l of degradation levels $(S = \bigcup_l s_l), l \in \mathbb{N}$. A membership function $\mu_{s_l}(V)$, $l = 1, \cdots, L$, characterizes each subset, where V is the level of degradation measured by the feature extracted from the vibration signal and L the number of degradation stages.
2) Let's consider the RUL as a fuzzy set Z defined on the time universe of run-to-failure of the bearings, and let's consider a finite number of fuzzy subsets z_l of $(Z = \bigcup_l z_l), l \in \mathbb{N}$. A membership function $\mu_{z_l}(\rho), l = 1, \cdots, L$, characterizes each subset, where ρ is the consumed lifetime feature corresponding to the degradation measured (Figure 3.3).

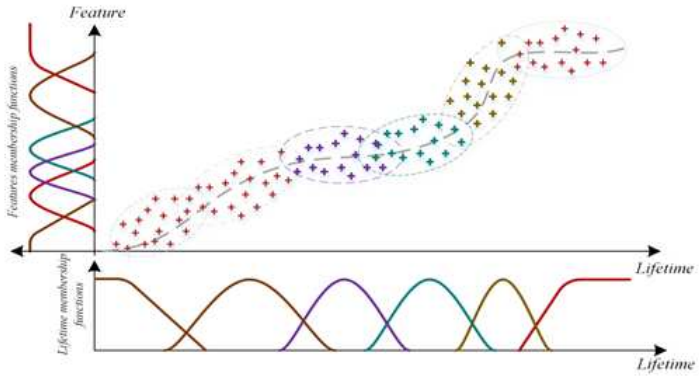


Figure 3.3 Relations graph of degradation feature and lifetime feature.

Definition

A clustering rule is an application that associates a degradation feature input data to one, and only one fuzzy subset of degradation and the corresponding lifetime.

Hypothesis: Degradation stage

We assume that degradation features can reveal the bearings degradation state to some extend with respect to the time features as well as the degradation stage.

Proposition 1

Based on the previous hypothesis, for each degradation stage, there exists a rule that associates a number of degradation features data to degradation state.

Proposition 2

Based on the previous hypothesis and proposition 1, a rule can associate a data of a fuzzy subset to only one degradation state.

Property 1

Assume two rules R_c and R_j $(c \neq j)$. If for the whole fuzzy set of premises, the application of the two rules yields the same results, then R_c is identical to R_j.

Property 2

The number of rules is equal to the number of degradation states.

Proof:

Let us denote L the number of degradation states and J the number of rules.

a) Assume $L < J$, then there exists at least one rule that associates the same data to at least two states, and so the proposition 2 is not fulfilled.
b) Assume $L > J$, and that and the proposition 1 and are fulfilled, then there exist at least two rules R_c and R_j $(c \neq j)$ that associates the same dataset to the same states. Then R_c is identical to R_j.

So, $L = J$.

3.1.2 T-S FIS modeling of bearings RUL

Fuzzy logic is an efficient theory to handle imprecision, which can take imprecise observations as inputs and achieve precise values as outputs [197].

FIS is one of the most famous applications of fuzzy logic and fuzzy sets theory [198]. The advantages of Fuzzy Inference Systems are that: on the one hand, they can handle information implicitly; on the other hand, they can describe nonlinear relationships between inputs and outputs [199].

FIS, as fuzzy models, is composed of the following blocks: fuzzy IF-THEN rules, the membership functions of the fuzzy sets, the inference operations on the rules, a fuzzification interface, a defuzzification interface. Generally, classic fuzzy systems refer to either Mamdani or T-S type fuzzy model. A Mamdani type fuzzy model was first introduced in [200]. In Mamdani type fuzzy models, the outputs of rules, which are fuzzy sets, are combined into a single fuzzy set using the aggregation. Then, the combined output fuzzy set is defuzzified to obtain a final crisp output value. In contrast, in T-S type fuzzy models, the crisp values of linear functions are used in the consequent part of each rule. Compared with the Mamdani type fuzzy models, the T-S FIS is more computationally efficient and suitable for mathematical analysis [166].

The model of bearing degradation process, depicted in Figure 3.2, is nonlinear. One efficient method to model the behavior of such a nonlinear system consists of identifying the operating regimes [201]. Thus, one can first establish local linear models by selecting operating points, then merging these local linear models to construct a nonlinear model as a succession of continuous piecewise linear models:

$$\hat{\rho}_k = \sum_{j=1}^{J} \bar{w}_j(V_k)\,(a_j V_k + b_j) \tag{3.3}$$

where $\bar{w}_j(V_k)$ is the activation function of a local linear continuous piece of the nonlinear model representing the j^{th} ($j=1,2...J$) operating regime, with the parameters [a_j, b_j].

From here, we suggest using fuzzy sets to identify the operating regimes. Thus, the nonlinear model will be represented by a T-S fuzzy inference system [202], where the rules are defined as it follows:

R_j: If V_k is A_j then the degradation is described by (y_j, r_j), $j = 1, 2, \cdots, J$

where $y_j = a_j V_k + b_j$ defines the j^{th} degradation continue piece-wise local linear model, $a_j = [a_{j,1}, a_{j,2}, \ldots, a_{j,I}]$ and b_j are the parameters of the j^{th} local linear model, A_j is the related fuzzy subset, and $r_j \in [0, 1]$ is the weight of j^{th} rule. The designer of the fuzzy system usually chooses the value of r_j. When such knowledge is not available, let $r_j = 1$; thus, there is no effect of r_j on the inference process of rules.

When the input of the model is multivariate, the antecedent part "V_k *is* A_j" can be presented as a logical combination of propositions with univariate fuzzy sub-subsets defined for individual entry $v_{k,i}$ of the vector V_k, usually in the following conjunctive form:

R_j: If $v_{k,1}$ is $A_{j,1}$ and $v_{k,2}$ is $A_{j,2}$ and $v_{k,3}$ is $A_{j,3}$ and… $v_{k,I}$ is $A_{j,I}$, then ($y_j = a_j V_k + b_j$, r_j) $j = 1, 2, \ldots, J$.

The degree of fulfillment of rules is defined by the product of membership degrees by input individual membership functions $\mu_{j,i}(v_{k,i})$ and the rule's weight r_j as in (3.4):

$$w_j(V_k) = r_j \prod_{i=1}^{I} \mu_{j,i}(v_{k,i}) \tag{3.4}$$

Rules aggregation allows obtaining the output (3.5):

$$\hat{\rho}_k = \frac{\sum_{j=1}^{J} w_j(V_k)\,(a_j V_k + b_j)}{\sum_{j=1}^{J} w_j(V_k)} \tag{3.5}$$

$$\bar{w}_j(V_k) = \frac{w_j(V_k)}{\sum_{j=1}^{J} w_j(V_k)} \tag{3.6}$$

The substitution of the normalized degree of rule's fulfillment $\bar{w}_j(X_k)$ (3.6) into (3.5) yields as (3.3).

In the next section, we introduce a method to identify the classic structure T-S FIS for bearings RUL estimation under the assumptions previously stated.

3.2 Fuzzy subtractive clustering

Clustering methods are the unsupervised methods that classify the observations, data items, or feature vectors in data space into clusters. According to the way of dividing the boundaries between clusters, there are two categories of clustering methods, i.e., crisp clustering methods and fuzzy clustering methods [203]. Crisp clustering methods are subject to the crisp boundaries between obtained clusters. While fuzzy clustering methods associate each object in the collection to each cluster with a certain degree.

Taking advantage of fuzzy logic, fuzzy clustering methods provide simple powerful techniques to handle imprecision and identify complex relationships existed in data space. Among the fuzzy clustering methods, there are two famous clustering techniques, i.e., fuzzy C-mean algorithm (FCM) and subtractive clustering algorithm [204]. FCM is by far the most popular objective function based fuzzy clustering algorithm, which is first developed in [205] and then improved in [206]. In [207], a simple and effective algorithm, called the mountain method, is proposed to estimate the number and the initial location of cluster centers. The mountain method is based on the grid of the data space, computing a potential value for each grid point. The value of the potential increases with the number of neighbors of the grid point. The grid point with the biggest potential value is the first cluster centroid. Each time a new cluster centroid is obtained, potential values of all grid points are reduced according to their distance from the cluster centroid. The next cluster centroid is the grid point with the biggest retained

potential value. This process of selecting a new cluster and reducing the potential values of all grid points repeats until the potential values of all grid points are less than a predefined threshold.

The significant drawback of the mountain method is that the computation load increases exponentially along with the dimension increasing of the problem, i.e., the number of grid points that must be evaluated increases exponentially along with the dimension [204]. In [208], the fuzzy subtractive clustering algorithm is proposed, an extension of the mountain method, to solve the computational problem. In fuzzy subtractive clustering, it considers each data point rather than each grid point. Thus, it eliminates the work of specifying grid resolutions to decrease the computation load.

According to the comparison in [204], the models established by using subtractive clustering usually are more accurate than those established by using the FCM algorithm. Different running of the FCM yields different results. In contrast, the fuzzy subtractive algorithm produces consistent results. Moreover, tuning the radii parameter of the fuzzy subtractive clustering is an efficient way to control the obtained number of clusters compared with FCM.

In this work, the fuzzy subtractive clustering is adopted because of its advantages aforementioned. A brief description of this method is provided in the sequel.

Considering a collection *G* of *K* data points $\dot{x}_k$ in an *I* dimensional space, the fuzzy subtractive clustering is implemented as follows.

Step 1, normalizing *G* in each dimension to be *X* with *K data points* x_k. Thus, *X* is bounded in a hypercube.

Step 2, calculating the potential measure (to be centroid) of all the data points in *X*, then selecting the data point with the biggest value of potential measure as the first cluster centroid. The first cluster centroid is denoted by x_1, and the value of its potential measure is denoted by η_1. The potential measure of data point x_k is defined in (3.7).

$$\eta_{k[j]} = \sum_{k'=1}^{K} e^{-\alpha\|x_k - x_{k'}\|^2} \tag{3.7}$$

where

$$\alpha = \frac{4}{r_a^2} \tag{3.8}$$

is the influence constant r_a is a positive constant. and *j* is the cluster centroid label.

Step 3, iteratively revising further the potential measure of all the data points in *X* by formula (3.9). In each sub-step, a data point is selected with the biggest value of a potential measure $\eta_{k[j]}$, which is denote by x_j, and the value of its potential measure is denoted by η_j.

$$\eta_{k[j]} = \eta_{k[j-1]} - \eta_{\bar{j}} e^{-\beta \|x_k - x_{\bar{j}}\|^2} \tag{3.9}$$

$$\beta = \frac{4}{r_b^2} \tag{3.10}$$

and the subtract constant r_b is a positive constant.

Step 4, Testing the stop criterion of the iteration. The following pseudo-code expresses the stop criterion.

---

if $\eta_{\bar{j}} > \bar{\varepsilon}\eta_1$,

Accept $x_{\bar{j}}$ as a cluster centroid and continue the clustering process.

else

if $\eta_{\bar{j}} < \underline{\varepsilon}\eta_1$,

reject $x_{\bar{j}}$ and end the clustering process.

else

If $\frac{d_{min}}{r_a} + \frac{\eta_j}{\eta_1} \geq 1$,

Accept $x_{\bar{j}}$ as a cluster centroid and continue the clustering process.

else

Reject $x_{\bar{j}}$ and set its corresponding potential value to 0.

Select the data point with the next biggest potential value as $x_{\bar{j}}$, its potential value as the $\eta_{\bar{j}}$. Retest.

end

end

end

---

where d_{min} denotes the shortest of the distances between $x_{\bar{j}}$ and all previously found cluster centroids. $\bar{\varepsilon}$ is the acceptance threshold, and $\underline{\varepsilon}$ is the rejection threshold.

Step 5, denormalizing $x_{\bar{j}}$, the obtained corresponding values are denoted by $\dot{x}_{\bar{j}}$ which are the cluster centroids of data space *G*.

3.3 Maximum-likelihood estimation of a multivariate mixture distribution

We assume the sample datasets *V* derive from a finite and countable mixture multivariate distribution, which is a mixture of multivariate normal distributions. Let θ_{R_j} denotes the j^{th} regime. j=1, 2, …, J is the label of regimes. We define the mixture distribution $f(V_k)$ by equations (3.11), (3.12), and (3.13).

$$f(V_k) = \sum_{j=1}^{J} \varsigma_j \, g_j(V_k; C_j, H_j) \tag{3.11}$$

$$\sum_{j=1}^{J} \varsigma_j = 1 \tag{3.12}$$

$$g_j(V_k; C_j, H_j) = \frac{1}{(2\pi)^{I/2}\left|H_j\right|^{1/2}} e^{(-\frac{1}{2}(V_k - C_j)H_j{}^{-1}(V_k - C_j)^T)} \tag{3.13}$$

$$P(\theta_{R_j}|V_k) = \frac{P(\theta_{R_j})P(V_k\left|\theta_{R_j}\right.)}{P(V_k)} = \frac{\varsigma_j g_j(V_k|C_j, H_j)}{f(V_k)} \tag{3.14}$$

where H_j and C_j are the covariance matrix and the centroids vector associated with the θ_{R_j} regime. The probability of membership $P(\theta_{R_j}|V_k)$ is defined in (3.14) [209], according to Bayesian rule.

The $f(V_k)$ can be considered as a function with ς_j, H_j, and C_j as its parameters.

The maximum-likelihood estimation of multivariate mixture distribution is an effective way of mixture distribution analysis [209]. Based on (3.11), (3.12) and (3.13), using the Lagrange multiplier method, the problem of multivariate mixture distribution parameters estimation with the maximum-likelihood is formulated as

$$\begin{aligned} \hat{\varsigma}_j &= \underset{\varsigma_j}{argmax}\Big(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\Big(\sum_{j=1}^{J}\varsigma_j - 1\Big)\Big) \\ &= \underset{\varsigma_j}{argmax}\Big(\sum_{k=1}^{K} ln \sum_{j=1}^{J}\varsigma_j \, g_j(V_k; C_j, H_j) - \lambda\Big(\sum_{j=1}^{J}\varsigma_j - 1\Big)\Big) \end{aligned} \tag{3.15}$$

$$\begin{aligned} \hat{\sigma}^2{}_{j,i} &= \underset{\sigma^2{}_{j,i}}{argmax}\Big(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\Big(\sum_{j=1}^{J}\varsigma_j - 1\Big)\Big) \\ &= \underset{\sigma^2{}_{j,i}}{argmax}\Big(\sum_{k=1}^{K} ln \sum_{j=1}^{J}\varsigma_j \, g_j(V_k; C_j, H_j) - \lambda\Big(\sum_{j=1}^{J}\varsigma_j - 1\Big)\Big) \end{aligned} \tag{3.16}$$

$$\begin{aligned} \hat{c}_{j,i} &= \underset{c_{j,i}}{argmax}\Big(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\Big(\sum_{j=1}^{J}\varsigma_j - 1\Big)\Big) \\ &= \underset{c_{j,i}}{argmax}\Big(\sum_{k=1}^{K} ln \sum_{j=1}^{J}\varsigma_j \, g_j(V_k; C_j, H_j) - \lambda\Big(\sum_{j=1}^{J}\varsigma_j - 1\Big)\Big) \end{aligned} \tag{3.17}$$

where $\lambda$is the Lagrange multiplier, $\hat{c}_{j,i}$and $\hat{\sigma}^2{}_{j,i}$ are the maximum-likelihood estimated mean and variance of the i^{th} feature of V associated with θ_{R_j}.

Since the mixture distribution $P(V_k)$ is a mixture of multivariate normal distributions, the function in the brackets of the right side of equations (3.15), (3.16) and (3.17) is convex. Hence, for example, when $P(\theta_{R_1})$ takes the maximum-likelihood value $\hat{P}(\theta_{R_1})$, the equation (3.18) is

ture. Similarly, when ${\sigma^2}_{1,1}$ takes the maximum-likelihood value ${\hat{\sigma}^2}_{1,1}$, the equation (3.19) is ture, when $c_{1,1}$takes the maximum-likelihood value $\hat{c}_{1,1}$, the equation (3.20) is ture.

$$\frac{\partial(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j \, g_j(V_k; C_j, H_j) - \lambda(\sum_{j=1}^{J} \varsigma_j - 1))}{\partial \varsigma_j} = 0 \tag{3.18}$$

$$\frac{\partial(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j \, g_j(V_k; C_j, H_j) - \lambda(\sum_{j=1}^{J} \varsigma_j - 1))}{\partial {\sigma^2}_{j,i}} = 0 \tag{3.19}$$

$$\frac{\partial(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j \, g_j(V_k; C_j, H_j) - \lambda(\sum_{j=1}^{J} \varsigma_j - 1))}{\partial c_{j,i}} = 0 \tag{3.20}$$

3.3.1 **Maximum-likelihood estimation of ς_j**

Substituting (3.14) into (3.18), (3.21) is obtained. Then, (3.22) is obtained by multiplying ς_j on both sides of the equation (3.21).

$$\sum_{k=1}^{K} \frac{g_1(V_k; C_j, H_j)}{f(V_k)} - \lambda = 0 \tag{3.21}$$

$$\varsigma_j * (\sum_{k=1}^{K} \frac{g_1(V_k; C_j, H_j)}{f(V_k)} - \lambda) = \varsigma_j * 0 \tag{3.22}$$

According to the formula (3.14), (3.23) can be obtained.

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) - \lambda \varsigma_j = 0 \tag{3.23}$$

Then, (3.24) is true.

$$\sum_{j=1}^{J} (\sum_{k=1}^{K} P(\theta_{R_j}|V_k) - \lambda \varsigma_j) = 0 \tag{3.24}$$

Hence,

$$\sum_{k=1}^{K} (\sum_{j=1}^{J} P(\theta_{R_j}|V_k)) - \lambda \sum_{j=1}^{J} \varsigma_j = 0 \tag{3.25}$$

Since $\sum_{j=1}^{J} P(\theta_{R_j}|V_k)$=1, and according to (3.12), then

$$K - \lambda = 0$$

$$\lambda = K \tag{3.26}$$

Substituting (3.26) into (3.23), (3.27) is obtained.

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) - K \varsigma_j = 0 \tag{3.27}$$

The solution of (3.27) is the maximum-likelihood value of ς_j as in (3.28).

$$\hat{\varsigma}_j = \frac{1}{K}\sum_{k=1}^{K} P(\theta_{R_j}|V_k) \tag{3.28}$$

3.3.2 Maximum-likelihood estimation of $\sigma^2{}_{j,i}$

Substituting (3.14) into (3.19), (3.29) is obtained.

$$\sum_{k=1}^{K} \frac{\varsigma_j}{f(V_k)} \frac{\partial g_j(V_k; C_j, H_j)}{\partial \sigma^2{}_{j,i}} = 0 \tag{3.29}$$

Changing the form of (3.29) to be (3.30), then, (3.31) is derived from (3.30).

$$\sum_{k=1}^{K} \frac{\varsigma_j g_j(V_k; C_j, H_j)}{f(V_k)} * \frac{1}{g_j(V_k; C_j, H_j)} * \frac{\partial g_j(V_k; C_j, H_j)}{\partial \sigma^2{}_{j,i}} = 0 \tag{3.30}$$

$$\sum_{k=1}^{K} \frac{\varsigma_j g_j(V_k; C_j, H_j)}{f(V_k)} * \frac{\partial ln g_j(V_k; C_j, H_j)}{\partial \sigma^2{}_{j,i}} = 0 \tag{3.31}$$

According to (3.14), (3.31) could be rewritten to be (3.32).

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) * \frac{\partial ln g_j(V_k; C_j, H_j)}{\partial \sigma^2{}_{j,i}} = 0 \tag{3.32}$$

In case of that, the variables in dataset V are linearly independent and under the condition of the branch distribution with an unequal covariance matrix, the expression (3.33) is obtained by substituting (3.13) into (3.32).

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) \left[-\frac{1}{2\sigma^2{}_{j,i}} + \frac{1}{2(\sigma^2{}_{j,i})^2}(v_{k,i} - c_{j,i})^2\right] = 0 \tag{3.33}$$

The expression (3.34) is obtained by multiplying $-2(\sigma^2{}_{j,i})^2$ on both sides of (3.33). Then, (3.35) is obtained.

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) \left[\sigma^2{}_{j,i} - (v_{k,i} - c_{j,i})^2\right] = 0 \tag{3.34}$$

$$\hat{\sigma}^2{}_{j,i} = \frac{\sum_{k=1}^{K} P(\theta_{R_j}|V_k)(v_{k,i} - c_{j,i})^2}{\sum_{k=1}^{K} P(\theta_{R_j}|V_k)} \tag{3.35}$$

3.3.3 Maximum-likelihood estimation of $c_{j,i}$

Substituting (3.14) into (3.20), (3.36) is obtained.

$$\sum_{k=1}^{K} \frac{\varsigma_j}{f(V_k)} \frac{\partial g_j(V_k; C_j, H_j)}{\partial c_{j,i}} = 0 \tag{3.36}$$

Changing the form of (3.36) to be (3.37), then, (3.38) is derived from (3.37).

$$\sum_{k=1}^{K} \frac{\varsigma_j g_j(V_k; C_j, H_j)}{f(V_k)} * \frac{1}{g_j(V_k; C_j, H_j)} * \frac{\partial g_j(V_k; C_j, H_j)}{\partial c_{j,i}} = 0 \tag{3.37}$$

$$\sum_{k=1}^{K} \frac{\varsigma_j g_j(V_k; C_j, H_j)}{f(V_k)} * \frac{\partial ln g_j(V_k; C_j, H_j)}{\partial c_{j,i}} = 0 \tag{3.38}$$

According to (3.14), (3.38) can be transformed into (3.39).

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) * \frac{\partial ln g_j(V_k; C_j, H_j)}{\partial c_{j,i}} = 0 \tag{3.39}$$

In case of that, the variables in dataset V are linearly independent and under the condition of the branch distribution with an unequal covariance matrix, then (3.40) is obtained by substituting (3.13) into (3.39).

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) \left[-\frac{1}{{\sigma^2}_{j,i}}(v_{k,i} - c_{j,i})\right] = 0 \tag{3.40}$$

Multiplying $-{\sigma^2}_{1,1}$ on both sides of (3.40), (3.41) is obtained. Then, (3.42) is obtained.

$$\sum_{k=1}^{K} P(\theta_{R_j}|V_k) (v_{k,i} - c_{j,i}) = 0 \tag{3.41}$$

$$\hat{c}_{j,i} = \frac{\sum_{k=1}^{K} P(\theta_{R_j}|V_k) v_{k,i}}{\sum_{k=1}^{K} P(\theta_{R_j}|V_k)} \tag{3.42}$$

3.4 Fuzzy inference system identification

We used training datasets obtained from a sample of bearings for identifying the T-S FIS model parameters. This task consists of determining the number of rules and the input and output parameters of each rule. Thus, the T-S FIS identification process consists of two steps:

First, based on the technique of fuzzy subtractive clustering [208], fuzzy clusters are identified in the sample space. The resulting clusters constitute fuzzy subsets characterizing particular behaviors of the input space used for feeding the T-S FIS-model. One then identifies the number J of rules and the parameters of input membership functions $\mu_{j_{,i}}(v_{k,i})$ for each rule.

Second, based on the mixture distribution analysis, we estimated the priori probabilities of fuzzy subsets and extracted time clusters. The extracted time clusters are associated with information characterizing particular degradation stages for a population of bearings and represent the projection of the input space onto the lifetime space. Furthermore, using the priori

probabilities obtained for the fuzzy subsets and the extracted time clusters, the weighted least square estimation is employed to identify the parameters of output membership functions of each fuzzy subset. Thus, the obtained parameters of output membership function embed more information associated with the population and lifetime of bearings. The proposed method for parameters of output membership function identification constitutes the main contribution of this research work described in subsection 3.5.2.

3.4.1 **The FIS-model input parameters identification**

The clusters obtained by the fuzzy clustering technique describe different regimes in the sample space. For each regime, a rule is defined in the T-S FIS model that specifies the output of the model based on the input space, which is considered as a local submodel of the system [165]. The form and the overlap of the input membership functions, identified by using the fuzzy clustering technique, could provide information about the nonlinear characteristic of the system[165].

The fuzzy subtractive clustering method used to identify the FIS model is characterized by its immune to data outliers, and low computational complexity, compared to existing complex FIS model identification methods under similar accuracy [208]. Hence, in this work, we use the technique of fuzzy subtractive clustering to identify the number J of rules and the input membership functions $\mu_{j,i}(v_{k,i})$. To build the training datasets for identifying the T-S FIS model, we use input V_k extracted from vibration signal data collected through periodic observations on a small number of bearings over an entire run-to-failure period.

The actual consumed useful lifetime ratio corresponding to k^{th} sample interval in the training datasets is denoted by ρ_k.

For training the model, we use K successive training data, all gathered in a $K \times (I+1)$ matrix G (3.43). Each row of the G matrix is corresponding to V_k in the training datasets. The first I columns correspond to the input, and the last column is the output of the training sample.

$$G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_K \end{bmatrix} = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,I} & \rho \\ v_{2,1} & v_{2,2} & \cdots & v_{2,I} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{K,1} & v_{K,2} & \cdots & v_{K,I} & \rho_K \end{bmatrix} \tag{3.43}$$

Each rule corresponds to a specific approximated linear behavior and can be associated with a cluster in the sample space. Then, we used the fuzzy subtractive clustering method [208] to identify the clusters based on the training samples in the rows of the matrix G. A rule is identified to each regime corresponding to each obtained cluster. Hence, the number J of obtained clusters is the number of the rules according to Property 2. The centroids of the clusters

obtained are memorized in the D matrix given in equation (3.44). The centroid of the j^{th} cluster in the input space constitutes the first I elements of the j^{th} row, while c_j^* is the centroid of the j^{th} cluster in the output space. The Gaussian function defined in equation (3.45) is highly employed in the literature to model the input membership functions in various application fields due to its invariance property under multiplication [210]. Therefore, we chose the Gaussian functions to model the inputs' membership functions.

To let (3.45) be consistent with the potentiality measurement of data point for being a cluster in the fuzzy subtractive clustering method [208], the parameter σ_i of the input membership function of the i^{th} input feature is defined in equation (3.46).

$$D = \begin{bmatrix} \mathcal{g}_1 \\ \mathcal{g}_2 \\ \vdots \\ \mathcal{g}_J \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2,} & \cdots & c_{1,I} & c_1 * \\ c_{2,1} & c_{2,2} & \cdots & c_{2,I} & c_2 * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{J,1} & c_{J,2} & \cdots & c_{J,I} & c_J * \end{bmatrix} \tag{3.44}$$

$$\mu_{j,i}(v_{k,i}) = e^{\frac{-(v_{k,i}-c_{j,i})^2}{2{\sigma_i}^2}} \tag{3.45}$$

$$\sigma_i = r_a(max(V_i) - min(V_i))/2\sqrt{2} \tag{3.46}$$

where j=1,2,..., J is the label of the rules, i=1,2..., I is the feature label of the input, V_i is the i^{th} input feature, r_a is the influence constant of fuzzy subtractive clustering.

The value of constant r_a of the fuzzy subtractive clustering influences the number of clusters. The bigger r_a causes fewer clusters. To tradeoffs between accuracy and calculation cost of the model execution, we chose r_a as 0.5. We follow the suggestions in [208] to fix the acceptance threshold to $\bar{\varepsilon}$=0.5 and rejection threshold to $\underline{\varepsilon}$=0.15. Additionally, we set the subtract constant r_b=1.25r_a for the i^{th} feature range in this work, which results in better performance of model compared with the r_b=1.5r_a suggested in [208].

3.4.2 The FIS-model output parameters identification

The parameters of output membership functions in each rule are identified using a training sample dataset. The sample datasets derive from the bearing degradation monitoring.

The structure and multi-positionality of faults lead to the complex evolution of bearings' degradation. Hence, we assume the sample datasets derive from a finite and countable mixture multivariate distribution$P(V_k)$, which is a mixture of multivariate normal distributions. Let θ_{R_j} denotes the j^{th} regime associated with the rule R_j of the T-S FIS, we define the mixture distribution $P(V_k)$ by equations (3.11) (3.12) (3.13).

According to the definition in formula (3.6), the normalized rule degree of fulfillments $\bar{w}_j(V_k)$ are with respect to the following conditions:

$$\bar{w}_j(V_k) \in [0,1], \forall j,k; \sum_{j=1}^{J} \bar{w}_j(V_k) = 1, \forall k; 0 < \sum_{k=1}^{K} \bar{w}_j(V_k) < K, \forall j.$$

Considering the definition of probability of membership in (3.14), $\bar{w}_j(V_k)$ can be considered as the estimated value of the probability of membership $P(\theta_{R_j}|V_k)$. Hence, $\bar{w}_{j,k}$ can be rewritten as $\hat{P}(\theta_{R_j}|V_k)$.

According to the formulas (3.28)(3.35)(3.42), we can write as follows:

$$\hat{P}(\theta_{R_j}) = \hat{\varsigma}_j = \frac{1}{K}\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k) = \frac{1}{K}\sum_{k=1}^{K} \bar{w}_j(V_k) \tag{3.47}$$

$$\hat{c}_{j,t} = \frac{\sum_{k=1}^{K} \tau_k \hat{P}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k)} = \frac{\sum_{k=1}^{K} \tau_k \bar{w}_j(V_k)}{\sum_{k=1}^{K} \bar{w}_j(V_k)} \tag{3.48}$$

$$\hat{\sigma}_{j,t}^2 = \frac{\sum_{k=1}^{K} (\tau_k - \hat{c}_{j,t})^2 \hat{P}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k)} = \frac{\sum_{k=1}^{K} (\tau_k - \hat{c}_{j,t})^2 \bar{w}_j(V_k)}{\sum_{k=1}^{K} \bar{w}_j(V_k)} \tag{3.49}$$

where τ_k is the time-coordinate of the input V_k associated with the corresponding bearing lifetime, $\hat{c}_{j,t}$ and $\hat{\sigma}_{j,t}^2$ are the maximum-likelihood estimated values of the time-coordinate centroid and the variance associated with θ_{R_j}. According to equations (3.48) (3.49), one can consider that the parameters $\hat{c}_{j,t}$ and $\hat{\sigma}_{j,t}^2$ describe the projection of the probability distribution of the V_k input onto the time-coordinate. Additionally, $\hat{\varsigma}_j$ could be considered as $\hat{P}(\theta_{R_j})$ which is the priori probability of the j^{th} regime of data space.

Following the approach in the previous section, we construct the input membership function for the time-coordinates in each rule of the T-S FIS model. Due to its invariance property under multiplication [210], we use the Gaussian type membership function for the time-coordinates, the same with the input space. Hence, the input membership function characterizing the time clusters is:

$$\hat{\mu}_{j,t}(\tau_k) = e^{\frac{-(\tau_k - \hat{c}_{j,t})^2}{2\hat{\sigma}_{j,t}^2}} \tag{3.50}$$

Now, we rewrite the formula (3.4) (3.6) (3.3),

$$\widetilde{w}_j(V_k) = \hat{P}(\theta_{R_j})\hat{\mu}_{j,t}(\tau_k) \prod_{i=1}^{I} \mu_{j,i}(v_{k,i}) = \hat{P}(\theta_{R_j})\hat{\mu}_{j,t}(\tau_k) w_j(V_k) \tag{3.51}$$

$$\bar{\widetilde{w}}_j(V_k) = \frac{\widetilde{w}_j(V_k)}{\sum_{j=1}^{J} \widetilde{w}_j(V_k)} \tag{3.52}$$

$$\hat{\rho}_k = \sum_{j=1}^{J} \bar{\widetilde{w}}_j(V_k)\,(a_j V_k + b_j) \tag{3.53}$$

In the left-hand side of the second equal sign of formula (3.51), the estimated maximum-likelihood $\hat{P}(\theta_{R_j})$ associated with the population is obtained by equation (3.47), also called the

priori probability, which can be considered working as the weight as r_j in (3.4). In parallel, $\hat{\mu}_{j,t}(\tau_k)$ in the formula (3.51) can be considered working as an individual membership degree of an extra dimension of input features. According to the right-hand side of the second equal sign of formula (3.51), $\widetilde{w}_j(V_k)$ is the weight $w_j(V_k)$ obtained by $\hat{P}(\theta_{R_j})\hat{\mu}_{j,t}(\tau_k)$. Comparing with $\bar{w}_j(V_k)$ in (3.6), $\bar{\tilde{w}}_j(V_k)$ in (3.52) brings extra information associated with population and lifetime of bearings.

For the sake of simplicity, let us define the following matrix and vectors:

the matrix:

$$\Phi = \begin{bmatrix} \bar{\tilde{w}}_1(V_1)V_1^T & \bar{\tilde{w}}_1(V_1) & \bar{\tilde{w}}_2(V_1)V_1^T & \bar{\tilde{w}}_2(V_1) & \cdots & \bar{\tilde{w}}_J(V_1)V_1^T & \bar{\tilde{w}}_J(V_1) \\ \bar{\tilde{w}}_1(V_2)V_2^T & \bar{\tilde{w}}_1(V_2) & \bar{\tilde{w}}_2(V_2)V_2^T & \bar{\tilde{w}}_2(V_2) & \cdots & \bar{\tilde{w}}_J(V_2)V_2^T & \bar{\tilde{w}}_J(V_2) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\tilde{w}}_1(V_K)V_K^T & \bar{\tilde{w}}_1(V_K) & \bar{\tilde{w}}_2(V_K)V_K^T & \bar{\tilde{w}}_2(V_K) & \cdots & \bar{\tilde{w}}_J(V_K)V_K^T & \bar{\tilde{w}}_J(V_K) \end{bmatrix}$$

the vectors:

$$\vec{\rho} = [\rho_1 \quad \rho_2 \quad \cdots \quad \rho_K]^T$$

$$\hat{\vec{\rho}} = [\hat{\rho}_1 \quad \hat{\rho}_2 \quad \cdots \quad \hat{\rho}_K]^T$$

$$\beta = [a_1 \quad b_1 \quad a_2 \quad b_2 \quad \cdots \quad a_J \quad b_J]^T.$$

Hence, for the entire training sample, formula (3.53) can be rewritten as follows:

$$\hat{\vec{\rho}} = \Phi\beta \tag{3.54}$$

For the training data, the left side and the first matrix of the right side are constants. The parameters are in the second matrix of right side. Thus, to determine the parameters of all the local models, it suffices to find the vector β, which minimizes $\|\vec{\rho} - \Phi\beta\|^2$, i.e., to solve the following problem using the linear least-squares estimation technique:

$$\hat{\beta} = \underset{\beta}{argmin}\|\vec{\rho} - \Phi\beta\|^2 \tag{3.55}$$

Since we use the maximum-likelihood estimated values to obtain the $\bar{\tilde{w}}_j(V_k)$ which is weighted, the linear least-squares estimation is weighted, and the solutions are embedded more information associated with population and lifetime of bearings, compared with using $\bar{w}_j(V_k)$.We use the Moore-Penrose inverse [211] minimizing (3.55) to obtain $\hat{\beta}$ as in (3.56).

$$\hat{\beta} = (\Phi^T\Phi)^{-1}\Phi^T\vec{\rho} \tag{3.56}$$

3.4.3 **Main steps for identifying the T-S FIS**

The identification process proposed in this chapter can be summarized hereafter.

Identifying the input parameters

1) Set the Gaussian input membership function of T-S FIS in equation (3.45).

2) Compute the matrix D matrix in equation (3.44).

3) Determine the rules number J and the parameter $c_{j,i}$ according to matrix D.

4) Compute the parameters σ_i using equation (3.46).

Identifying the output parameters

1) According to (3.47), estimate the priori probability of a given degradation stage $\hat{P}(\theta_{R_j})$.

2) Project the feature space clusters on the time-space, with $\hat{\mu}_{j,.}$ obtained using (3.50).

3) Estimate $\tilde{\bar{w}}_j(V_k)$ using (3.52).

4) Calculate the output parameters a_j and b_j using (3.56).

Additionally, setting the value of r_j to 1. Lastly, the parameters r_j, J, $c_{j,i}$, σ_i, a_j and b_j allow building the T-S FIS model.

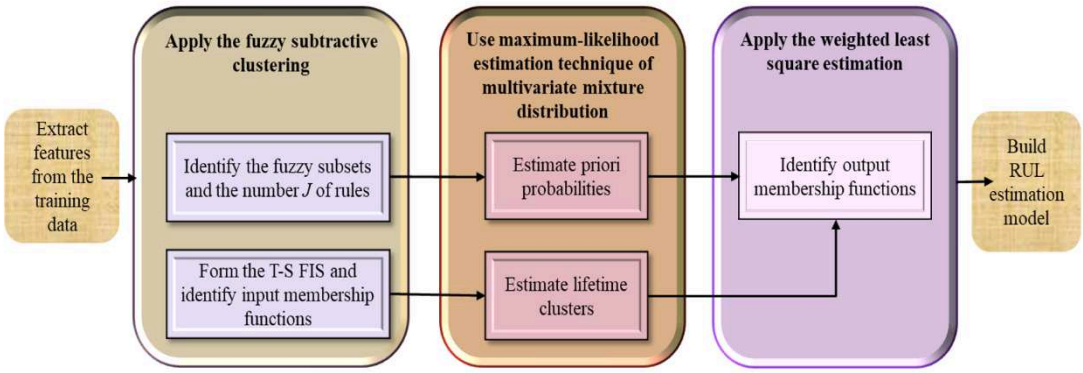


Figure 3.4 The flow diagram of the proposed algorithm for the T-S FIS model identification.

3.5 Summary

In this chapter, we proposed a new method to identify effectively a standard T-S FIS designed for estimating bearings RUL, using historical data over an entire run-to-failure cycle collected for a small population of identical bearings. To offset the impact of the lack of data, we proposed a method based on the maximum likelihood estimation of mixture distribution analysis.

During the T-S FIS model identification by the proposed method, using the priori probabilities obtained for the fuzzy subsets and the extracted time clusters, the weighted least square estimation has allowed identifying the parameters of output membership functions of each fuzzy subset. Thus, the obtained parameters of the output membership function embed information associated with the population due to the priori probabilities and with the lifetime of bearings due to the time clusters.

It makes sense that the proposed method can extract and use extra useful information from a small size dataset for model identification. In the next chapter, we propose an appropriate method of the FIS model parameters adaption, starting from the newly collected small size dataset.

Chapter 4 - Fuzzy model tuning based on small-size samples mixture distribution analysis for bearings remaining useful life estimation

The research presented in this chapter aims to find an effective way to tune a classic T-S FIS with simple structure to obtain a more accurate model for estimating bearings RUL using the vibration signals observed over a complete run-to-failure cycle from a small number of identical bearings operating under identical load. To offset the drawback of small size sample for model identification, in this chapter we propose a mixture distribution analysis based fuzzy model tuning approach to update the parameters of the T-S FIS model from the initial model by using small size additional training data sets to obtain a more accurate model. First, the input parameters of T-S FIS model are updated by using an iterative maximum likelihood estimation of mixture distribution analysis based method which is in this chapter. Then, the output parameters of T-S FIS model are updated by using another maximum likelihood estimation of the mixture distribution analysis based method which is proposed in this chapter by using the symmetrical Kaczmarz method. Finally, the tuned T-S FIS model is formed by using the updated parameters.

4.1 Problem statement

As we mentioned in the previous chapter, let $V_k=[v_{k,1},\ldots,v_{k,I}]^T$ denotes the input vector of the model for bearings RUL estimation, which is constituted by the features or indicators extracted from the vibration signals of k^{th} sample interval during bearings monitoring, where k=1,2,...,K, and i=1,2,...,I is the variable label.. The corresponding output of the model is the estimated consumed useful life ratio denoted by $\hat{\rho}_k$. The model for estimating bearings RUL is summarized in Figure 3.2, which is a classic T-S FIS with a simple structure.

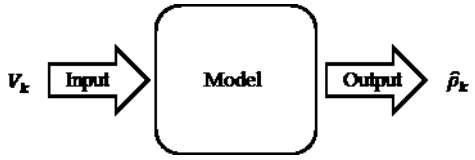


Figure 4.1 The model for RUL estimation.

Based on the indicators extracted from the vibration signal observed during the complete run to failure lifecycle of a small number of identical bearings under identical load, this work presented in this chapter aims to find an effective way to tune a T-S FIS model and obtain a more accurate tuned model for estimating bearings RUL, which appears inside the dot-dash box in Figure 4.2. The initial model is identified by the method presented in the previous chapter, the identification process of which appears inside the dashed box in Figure 4.2.

The sample size seriously affects the performance of the algorithms, especially in the case of a small-size sample [212]. Generally, to tune models with small-size samples could cause the developing process to be oscillating and non-robust unless the modeling object really lacks diversity. For this work, the modeling objects are degradation processes of bearings which are diverse. Hence, the influence of the drawback of small-size samples for model identification in this work is more significant.

We assume that to offset the influence of the drawback of small size sample to obtain a model with good performance for estimating bearings RUL, a model could be tuned to be more accurate by updating parameters of the model based on the small size additional training date set. In this work, the additional training data sets are derived from the vibration signal observed during the complete run to failure lifecycle of a small number of identical bearings under identical load.

In the T-S FIS model, the parameters include input parameters, output parameters, and the rules number. Due to the small size of the additional training date set, another assumption in this

chapter is that the tuning process involves only the parameters of input and output membership functions. Therefore, the number of rules in the tuned T-S FIS model is the same as it in the initial T-S FIS model.

Main contributions of this research are:

(1) Iterative maximum likelihood estimation of a mixture distribution analysis based method is proposed which can offset the drawback of the small size sample to the model tuning process. By the proposed method, the input parameters of the T-S FIS model are tuned based on the values of the input parameters in the initial model;

(2) Combined with the symmetrical Kaczmarz method, a maximum likelihood estimation of a mixture distribution analysis based method is proposed which also can offset the drawback of the small size sample to model the tuning process. By the proposed method, the output parameters of the T-S FIS model are tuned based on the values of the output parameters in the initial model.

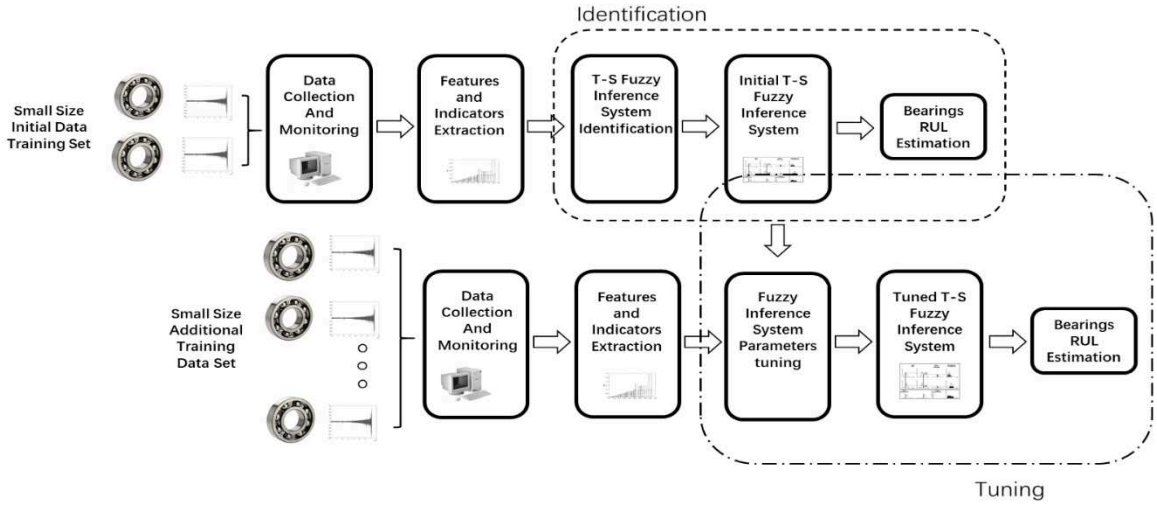


Figure 4.2 The tuning approach of T-S FIS model in this research.

4.2 T-S FIS model tuning method

As described in chapter 3, the parameters of a T-S FIS model include the input parameters $c_{j,i}$, σ_i, the output parameters a_j, b_j and the number of rules *J*.

In this work, the tuning process refers to parameters $c_{j,i}$, σ_i, a_j, b_j except for the number of rules *J*. Note that the number of rules *J* in the initial, and tuned T-S FIS model are the same.

Let $c_{j,i}$, σ_i, a_j, b_j denote the parameters in the initial T-S FIS model. Let $\sigma_i^{'}$, $c_{j,i}^{'}$, $a_j^{'}$, $b_j^{'}$ denote the parameters of the tuned T-S FIS model. The input of the additional training data set based

on which the tuned T-S FIS model is obtained is denoted by $V_k = [v_{k,1} \dots v_{k,I}]^T$. The output of the additional training data set is denoted ρ_k.

The process of tuning the initial T-S FIS model to the tuned T-S FIS model can be summarized in three steps.

Step 1: extracting the input parameters $c_{j,i}$, σ_i, the output parameters a_j, b_j, and the number of rules *J* from the initial T-S FIS model identified as proposed in the previous chapter.

Step 2: to offset the instability of the model tuning process caused by small size additional training set, the parameter $\sigma_i^{'}$ is obtained by using values of σ_i to preserve some of the characteristics of the initial model. Then, based on an iterative maximum likelihood estimation method proposed in the next section of this chapter, a set of the potential values of $c_{j,i}^{'}$ is obtained by using values of $c_{j,i}$. A validity measure method proposed in the next subsection of this chapter yields the suitable value of $c_{j,i}^{'}$ is selected.

Step 3: based on the mixture distribution analysis, we estimated the priori probabilities of fuzzy subsets and extracted time clusters, which correspond to the different bearing degradation stages. The extracted time clusters characterize degradation stages for a population of bearings and represent the projection of the input features space onto the bearings lifetime space. Furthermore, based on the priori probabilities obtained for the fuzzy subsets and the extracted time clusters, the symmetric Kaczmarz algorithm [213] is used to obtain the values of the output parameters $a_j^{'}$ and $b_j^{'}$ by using the values of a_j, b_j. Thus, the obtained parameters of the output membership function do not only carry more information associated with the bearings population and their lifetime but also offset the instability of the model tuning process caused by small size additional training set.

4.2.1 **Input parameters tuning**

Validity measure of fuzzy clustering based on cluster compactness and separation

The input parameters of the initial T-S FIS model are obtained by the fuzzy subtractive clustering method. Therefore, during the tuning process, the input parameters can be evaluated by the validity measure for fuzzy clustering.

Clustering methods are mostly unsupervised processes. Since the results of identical clustering algorithms on the same data set could be diverse [214]. A partition generated by a clustering algorithm does not always correspond to the actual underlying structure of data sets.

Cluster validity measures the quality of clustering results, which evaluates and compares the whole partitions resulting from different or the same clustering algorithm under different parameters. In general, there are mainly three categories of cluster validity methods, the methods measuring partition validity by evaluating the properties of the crisp structure respectively (e.g., [215] [216]), fuzzy memberships (e.g., [217] [218]) and both of the fuzzy memberships and structure (e.g., [219][220]) imposed on the data by the clustering algorithm [221].

In [221], to comprehensively measuring the partition validity in the case of fuzzy clustering, a cluster validity index was proposed that accounts for both the properties of fuzzy memberships and the data, which is based on compactness and separation. It is used to guide a (re)clustering process toward improved partition with an equal number of clusters. Compactness refers to the variation or spread manifested by the elements that belong to the same cluster. And separation represents the isolation of clusters from one another. The proposed clusters' validity index is the ratio of their compactness to their separation. To be more suitable for our work, we proposed a modified version of the clusters' validity index, as defined in (4.1).

$$Y = \sum_{j=1}^{J} \sum_{g=1}^{J} \frac{\sum_{i=1}^{I} (\sigma_i)^2}{\sum_{i=1}^{I} \left\| c_{j,i} - c_{g,i} \right\|^2} \tag{4.1}$$

where $g \neq j$. At the right side in equation (4.1), the numerator part is associated with the compactness of clusters, the denominator part is associated with the separation of clusters. The more compact the clusters are, the smaller the numerator part is. A large denominator part indicates the clusters are separated. Moreover, the value of *Y,* which is the ratio of compactness to separation, is associated with the discrepancy between clusters. The more diverse the clusters are, the smaller *Y* is. When the value of *Y* is less than 1, the center of each cluster is approximately out of the area of 68% confidence interval to other clusters in the case of Gaussian type fuzzy memberships.

Tuning the parameter $\boldsymbol{\sigma_i}$

The input of the T-S FIS model is fuzzified through input membership functions. The parameter σ_i determines the range of strong influence of Gaussian functions in data space [222]. In the case of the same number of input membership functions, too big value of $\boldsymbol{\sigma_i}$ will cause that the scope of influence of input membership functions goes beyond the scope of input data space, too small value of σ_i will cause that the scope of influence neglect part of the input data space.

According to the equation (3.46), a parameter $\sigma_i^{\#}$ associated with the additional data set V is defined in (4.2).

$$\sigma_i^{\#} = r_a(max(V_i) - min(V_i))/2\sqrt{2} \tag{4.2}$$

where V_i is the i^{th} variable of input.

To offset the instability of model tuning process due to the small-size of the additional training set, the parameter $\sigma_i^{\#}$ is proportionately fused with σ_i by (4.3) to obtain the new parameter σ_i', which is the parameter in the tuned model.

$$\sigma_i' = (\sigma_i + \frac{s'}{s} * \sigma_i^{\#})/(1 + \frac{s'}{s}) \tag{4.3}$$

where s' is the number of bearings in additional training data set, and s is the number of bearings in the initial training data set. Formula (4.3) is a modified version of that in [223] which simply fuses the two parameters by averaging.

Tuning the parameter $c_{j,i}$

Due to the complexity of the bearing degradation process, we assume the sample datasets V are derived from a finite and countable mixture multivariate distribution $f(V_k)$, which is a mixture of multivariate normal distributions $g_j(V_k; C_j, H_j)$ by taking proportions ς_j. Each multivariate normal distribution is associated with a specific regime θ_{R_j} in the data space. We define the mixture distribution density function $f(V_k)$ by equations (4.4)(4.5)(4.6).

$$f(V_k) = \sum_{j=1}^{J} \varsigma_j \, g_j(V_k; C_j, H_j) \tag{4.4}$$

$$\sum_{j=1}^{J} \varsigma_j = 1 \tag{4.5}$$

$$g_j(V_k; C_j, H_j) = \frac{1}{(2\pi)^{I/2}\left|H_j\right|^{1/2}} e^{(-\frac{1}{2}(V_k - C_j)H_j^{-1}(V_k - C_j)^T)} \tag{4.6}$$

$$P(\theta_{R_j}|V_k) = \frac{P(\theta_{R_j})P(V_k\left|\theta_{R_j}\right.)}{P(V_k)} = \frac{\varsigma_j g_j(V_k|C_j, H_j)}{f(V_k)} \tag{4.7}$$

where θ_{R_j} denotes the j^{th} regime associated with the j^{th} branch distribution, H_j and C_j are the covariance matrix and the mean vector associated with the regime θ_{R_j}. The probability of membership $P(\theta_{R_j}|V_k)$ is defined in (4.7) [209], according to Bayesian rule.

The $f(V_k)$ is a function that depends on parameters ς_j, H_j, and C_j.

The maximum-likelihood estimation of multivariate mixture distribution is an effective way of mixture distribution analysis [209]. Based on (4.4)(4.5)(4.6)(4.7) and using the Lagrange

multiplier method, the problem of multivariate mixture distribution parameters estimation with the maximum-likelihood is formulated as in (4.8).

$$\begin{aligned}\hat{c}_{j,i} &= \underset{c_{j,i}}{argmax}\left(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right) \\ &= argmax\left(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j\, g_j(V_k; C_j, H_j) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right)\end{aligned} \tag{4.8}$$

where λ is the Lagrange multiplier, $\hat{c}_{j,i}$ is the maximum-likelihood estimated mean of the i^{th} feature of V associated with θ_{R_j}, which is an element of C_j.

To numerically solve maximum likelihood equation (4.8), an iterative method could be used to assume a trial solution and derive linear equations for small additive corrections to obtain the solution iteratively [224]. According to the approximation proposed in [209] for obtaining the solution of the maximum likelihood equation (4.8), the parameter $\hat{c}_{j,i}$ can be iteratively estimated by (4.9) [209].

$$\hat{c}_{j,i}{}^{[l]} = \frac{\sum_{k^{\#}=1}^{K^{\#}} v_{k,i} \hat{P}^{[l-1]}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}^{[l-1]}(\theta_{R_j}|V_k)} \tag{4.9}$$

where l =1,2,… L is the label of the iterative steps. To offset the instability of the model tuning process due to the small size of the additional training data set, the parameter $\hat{c}_{j,i}{}^{[l]}$ is fused with $c_{j,i}$ as (4.10) to obtain $\tilde{c}_{j,i}{}^{[l]}$ [223]. The parameters in (4.10) can all be considered as the parameters of Gaussian functions, which are used as the fuzzy membership functions. The Gaussian functions, such kind of bell shape functions with bigger standard deviation have a bigger range of strong influence in data space [222]. From formula (4.10), it can make sense that the Gaussian membership function with a bigger standard deviation has greater influence in the fusing process.

$$\tilde{c}_{j,i}{}^{[l]} = (c_{j,i}\sigma_i + \hat{c}_{j,i}{}^{[l]}\sigma_i^{\#})/(\sigma_i + \sigma_i^{\#}) \tag{4.10}$$

Rewriting formula (3.45) with additional training data set in an iterative way yields:

$$\mu^{[l]}{}_{j,i}(v_{k,i}) = e^{\frac{-(v_{k,i}-\tilde{c}_{j,i}{}^{[l]})^2}{2(\tilde{\sigma}_i^{[l]})^2}} \tag{4.11}$$

$$\mu^{[l]}{}_{j}(V_k) = \prod_{i=1}^{I} \mu^{[l]}{}_{j,i}(v_{k,i}) \tag{4.12}$$

$$\bar{\mu}_j{}^{[l]}(V_k) = \frac{\mu^{[l]}{}_{j}(V_k)}{\sum_{j=1}^{J} \mu^{[l]}{}_{j}(V_k)} \tag{4.13}$$

According to the definition in formula (4.13), $\bar{\mu}_j{}^{[l]}(V_k)$ fulfills the following conditions:

$$\bar{\mu}_j{}^{[l]}(V_k) \in [0,1], \forall j,k,i; \sum_{j=1}^{J} \bar{\mu}_j{}^{[l]}(V_k) = 1, \forall k,i; 0 < \sum_{k=1}^{K} \bar{\mu}_j{}^{[l]}(V_k) < N, \forall j,i.$$

Considering the definition of probability of membership [209], $\bar{\mu}_j{}^{[l-1]}(V_k)$ is the estimated value of the probability of membership $P(\theta_{R_j}|V_k)$. Hence, $\bar{\mu}_j{}^{[l]}(V_k)$ could be rewritten as $\hat{P}^{[l]}(\theta_{R_j}|V_k)$.

According to the formula (4.9) and (4.13), we can write:

$$\hat{c}_{j,i}{}^{[l]} = \frac{\sum_{k=1}^{K} v_{k,i}\hat{P}^{[l-1]}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}^{[l-1]}(\theta_{R_j}|V_k)} = \frac{\sum_{k=1}^{K} v_{k,i}\bar{\mu}_j{}^{[l-1]}(V_k)}{\sum_{k=1}^{K} \bar{\mu}_j{}^{[l-1]}(V_k)} \tag{4.14}$$

In the above iterative process, when l=1, $\tilde{c}_{j,i}{}^{[l-1]} = c_{j,i}$, $\tilde{\sigma}_i^{[l-1]} = \sigma_i$, and when $l \neq 1$, $\tilde{\sigma}_i^{[l-1]} = \sigma_i^{\#}$, $\tilde{c}_{j,i}{}^{[l-1]}$ is obtained by (4.10). At the first step, it can make sense that $\hat{c}_{j,i}{}^{[1]}$ is to lead the characteristics of the initial model into the iterative process to offset the oscillation due to the small size of the additional training data set.

Considering that the number of clusters in each step is consistent, the validity measurement for evaluating the partitions with the same number of clusters in each step is proposed in (4.1). To be consistent with the process in this section, equation (4.1) could be rewritten as in (4.15).

$$Y^{[l]} = \sum_{j=1}^{J}\sum_{d=1}^{J} \frac{\sum_{i=1}^{I}(\sigma_i')^2}{\sum_{i=1}^{I}\left\|\tilde{c}_{j,i}{}^{[l]} - \tilde{c}_{d,i}{}^{[l]}\right\|^2} \tag{4.15}$$

where $d \neq j$.

Through predefined parameter $Maxiter$ which is the number of iterative steps, the $\tilde{c}_{j,i}{}^{[l]}$ of the step l corresponding to the smallest value of $Y^{[l]}$ is used as the value of $c_{j,i}^{'}$. Along with the iterative steps, the value of $Y^{[l]}$ is roughly convergent.

Main steps for input parameters tuning

The input parameters tuning process proposed in this chapter can be summarized hereafter.

Tuning the parameter σ_i

1) Calculate the parameter $\sigma_i^{\#}$ based on the additional training data set by formula (4.2).

2) fuse σ_i with $\sigma_i^{\#}$ to obtain σ_i' by formula (4.3).

Tuning the parameter $c_{j,i}$

1) Initialize l=1, set parameter $Maxiter$.

2) Using $\tilde{c}_{j,i}{}^{[l-1]} = c_{j,i}$, $\tilde{\sigma}_i^{[l-1]} = \sigma_i$, to obtain $\widehat{c}_{j,i}{}^{[l]}$ based on the iterative maximum likelihood estimation method by formula (4.14).

3) fuse $\widehat{c_{j,i}}^{[l]}$ with $c_{j,i}$ to obtain $\tilde{c}_{j,i}{}^{[l]}$ by formula (4.10).

4) Calculate validity measurement $Y^{[l]}$ by formula (4.15).

5) $l = l + 1$.

6) While $l \leq Maxiter$

Using $\tilde{c}_{j,i}{}^{[l-1]}$and $\tilde{\sigma}_i^{[l-1]} = \sigma_i^{\#}$, to calculate $\widehat{c_{j,i}}^{[l]}$ based on the iterative maximum likelihood estimation method by formula (4.14);

fuse $\widehat{c_{j,i}}^{[l]}$with $c_{j,i}$ to obtain $\tilde{c}_{j,i}{}^{[l]}$ by formula (4.10);

Calculate validity measurement $Y^{[l]}$ by formula (4.15);

$l = l + 1$;

End (While)

Select the $\tilde{c}_{j,i}{}^{[l]}$ with a minimum value of $Y^{[l]}$ as the value of $c'_{j,i}$.

Thus, one obtains the input parameters $\sigma_i{}'$ and $c'_{j,i}$ of the tuned T-S FIS model.

4.2.2 **Output parameters tuning**

Tuning parameters a_j and b_j

Based on (4.4)(4.5)(4.6)(4.7) and using the Lagrange multiplier method, the problem of multivariate mixture distribution parameters estimation with the maximum-likelihood formulated as it follows:

$$\hat{\varsigma}_j = \underset{\varsigma_j}{argmax}\left(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right) \tag{4.16}$$

$$= \underset{\varsigma_j}{argmax}\left(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j\, g_j(V_k; C_j, H_j) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right)$$

$$\hat{\sigma}^2{}_{j,i} = \underset{\sigma^2{}_{j,i}}{argmax}\left(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right) \tag{4.17}$$

$$= \underset{\sigma^2{}_{j,i}}{argmax}\left(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j\, g_j(V_k; C_j, H_j) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right)$$

$$\hat{c}_{j,i} = \underset{c_{j,i}}{argmax}\left(\sum_{k=1}^{K} ln\, f\,(V_k) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right) \tag{4.18}$$

$$= \underset{c_{j,i}}{argmax}\left(\sum_{k=1}^{K} ln \sum_{j=1}^{J} \varsigma_j\, g_j(V_k; C_j, H_j) - \lambda\left(\sum_{j=1}^{J} \varsigma_j - 1\right)\right)$$

Under the condition of the branch distributions with unequal covariance matrices, according to the derivations on (3.15)(3.16)(3.17) in [209], the maximum-likelihood estimated values of

parameters in a mixture of multivariate normal distributions can be obtained without an iterative process as follows:

$$\hat{\varsigma}_j = \frac{1}{K}\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k) \tag{4.19}$$

$$\hat{c}_{j,i} = \frac{\sum_{k=1}^{K} v_{k,i}\hat{P}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k)} \tag{4.20}$$

$$\hat{\sigma}_{j,i}^2 = \frac{\sum_{k=1}^{K} (v_{k,i} - \hat{c}_{j,i})^2\hat{P}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k)} \tag{4.21}$$

where $\hat{c}_{j,i}$ and $\hat{\sigma}_{j,i}$ are the maximum-likelihood estimated mean and variance of i^{th} variable of V associated with θ_{R_j}.

According to the definition in formula (3.6), the normalized rule degree of fulfillments $_{\bar{w}_j(V_k)}$ are respecting to the following conditions:

$$\bar{w}_j(V_k) \in [0,1], \forall j,k; \sum_{j=1}^{J} \bar{w}_j(V_k) = 1, \forall k; 0 < \sum_{k=1}^{K} \bar{w}_j(V_k) < K, \forall j.$$

Considering the definition of the probability of membership in [209], $\bar{w}_j(V_k)$ can be considered as the estimated value of the probability of membership $P(\theta_{R_j}|V_k)$. Hence, $\bar{w}_j(V_k)$ can be rewritten as $\hat{P}(\theta_{R_j}|V_k)$.

According to the formula (4.19)(4.20)(4.21), we can write:

$$\hat{P}(\theta_{R_j}) = \hat{\varsigma}_j = \frac{1}{K}\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k) = \frac{1}{K}\sum_{k=1}^{K} \bar{w}_j(V_k) \tag{4.22}$$

$$\hat{c}_{j,t} = \frac{\sum_{k=1}^{K} \tau_k\hat{P}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k)} \tag{4.23}$$

$$\hat{\sigma}_{j,t}^2 = \frac{\sum_{k=1}^{K} (\tau_k - \hat{c}_{j,t})^2\hat{P}(\theta_{R_j}|V_k)}{\sum_{k=1}^{K} \hat{P}(\theta_{R_j}|V_k)} \tag{4.24}$$

where τ_k is the time-coordinate of the input V_k in relation to the corresponding bearing lifetime, $\hat{c}_{j,t}$ and $\hat{\sigma}_{j,t}^2$ are the maximum-likelihood estimated values of the time-coordinate centroid and variance associated with θ_{R_j}. Note that the parameters $\hat{c}_{j,t}$ and $\hat{\sigma}_{j,t}^2$ represent the projection of the probability distribution of the input V_k on the time-coordinates.

Following the approach in chapter 3, we can construct the input membership function for the time-coordinates in each rule of the T-S FIS model. Due to its property of being invariant under multiplication [210], we use the Gaussian type membership function for the time-coordinates the same with the input membership function for input V_k. Hence, the input membership function of the time-coordinates clusters is:

$$\mu_{j,t}(\tau_k) = e^{\frac{-(\tau_k - \hat{c}_{j,t})^2}{2\hat{\sigma}_{j,t}^2}} \tag{4.25}$$

Now, we rewrite the formula (3.28)(3.35)(3.42):

$$\widetilde{w}_j(V_k) = \hat{\varsigma}_j \mu_{j,t}(\tau_k) \prod_{i=1}^{I} \mu_{j,i}(v_{k,i}) = \hat{P}(\theta_{R_j}) \mu_{j,t}(\tau_k) w_j(V_k) \tag{4.26}$$

$$\bar{\widetilde{w}}_j(V_k) = \frac{\widetilde{w}_j(V_k)}{\sum_{j=1}^{J} \widetilde{w}_j(V_k)} \tag{4.27}$$

$$\hat{\rho}_k = \sum_{j=1}^{J} \bar{\widetilde{w}}_j(V_k) \, (a'_j V_k + b'_j) \tag{4.28}$$

On the right side of the second equal sign of equation (4.26), the maximum-likelihood estimated $\hat{P}(\theta_{R_j})$ is obtained by equation (4.22), also called the priori probability of fuzzy subset, can be considered working as the weight as r_j in (3.4). $\mu_{j,t}(\tau_k)$ in equation (4.26) can be considered as an individual membership degree of an extra dimension of features.

According to the right-hand side of the second equal sign of equation (4.26), $\widetilde{w}_j(V_k)$ is $w_j(V_k)$ weighted by $\hat{P}(\theta_{R_j}) \mu_{j,t}(\tau_k)$. Comparing with $\bar{w}_j(V_k)$ in (3.4), $\bar{\widetilde{w}}_j(V_k)$ in (4.27) brings extra information about the population and lifetime of bearings.

For the sake of simplicity, let's define the following matrix and vectors:

$$\overrightarrow{\Phi'} = \begin{bmatrix} \bar{\widetilde{w}}_1(V_1)V_1^T & \bar{\widetilde{w}}_1(V_1) & \bar{\widetilde{w}}_2(V_1)V_1^T & \bar{\widetilde{w}}_2(V_1) & \cdots & \bar{\widetilde{w}}_J(V_1)V_1^T & \bar{\widetilde{w}}_J(V_1) \\ \bar{\widetilde{w}}_1(V_2)V_2^T & \bar{\widetilde{w}}_1(V_2) & \bar{\widetilde{w}}_2(V_2)V_2^T & \bar{\widetilde{w}}_2(V_2) & \cdots & \bar{\widetilde{w}}_J(V_2)V_2^T & \bar{\widetilde{w}}_J(V_2) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\widetilde{w}}_1(V_K)V_K^T & \bar{\widetilde{w}}_1(V_K) & \bar{\widetilde{w}}_2(V_K)V_K^T & \bar{\widetilde{w}}_2(V_K) & \cdots & \bar{\widetilde{w}}_J(V_K)V_K^T & \bar{\widetilde{w}}_J(V_K) \end{bmatrix} = [\Phi_1' \quad \Phi_2' \quad \cdots \quad \Phi_K']^T$$

$\vec{\rho} = [\rho_1 \quad \rho_2 \quad \cdots \quad \rho_K]^T$

$\beta' = [a_1' \quad b_1' \quad a_2' \quad b_2' \quad \cdots \quad a_J' \quad b_J']^T$.

Hence, for the entire training sample, according to equation (4.28), a linear system problem can be defined as follows:

$$\vec{\rho} = \overrightarrow{\Phi'} \beta' \tag{4.29}$$

For the training data, the left side and the first matrix of the right side are constants. The parameters are in the vector β'.

Since we use the maximum-likelihood estimated values to obtain the $\bar{\widetilde{w}}_j(V_k)$ which is weighted, the approximate solutions of the linear equations system problem in (4.29) embed more information about the population and lifetime of bearings, compared with using $\bar{w}_j(V_k)$.

In this work, the symmetrical Kaczmarz method is employed to solve the linear system problem in an iterative way [225]. The Kaczmarz algorithm is initially introduced in [213], which is an iterative row action algorithm that computes approximate solutions to a consistent and

overdetermined linear system such as in (4.29). In order to reduce the influence of the order of rows in the matrix $\vec{\Phi}$ on the results, the symmetric Kaczmarz algorithm is used.

The process of solving the linear system problem by the symmetric Kaczmarz algorithm could be described briefly by the pseudo code as follow:

Initialize $\beta^{[1]} = [a_1 \quad b_1 \quad a_2 \quad b_2 \quad \cdots \quad a_J \quad b_J]^T$, l_0=1;

While $l_0 \leq Maxiter1$

For i_0=1:2K

(let $u_1, u_2 \ldots, u_{2K}$= 1,2…, K-1, K, K, K-1, K-2…, 1.)

$\beta^{[i_0+1]} = \beta^{[i_0]} + \delta \frac{\rho_{u_{i_0}} - \Phi_{u_{i_0}}^T \beta^{[i_0]}}{\left\| \Phi_{u_{i_0}} \right\|^2} \Phi_{u_{i_0}}$, where δ is the relax parameter.

end

$l_0 = l_0 + 1$;

$\beta^{[1]} = \beta^{[2K]}$;

end

Let $\beta' = \beta^{[1]}$.

The initial values of $\beta^{[1]}$ are set as $\beta^{[1]} = [a_1 \quad b_1 \quad a_2 \quad b_2 \quad \cdots \quad a_J \quad b_J]^T$. Thus, the results naturally are embedded the information from both of the initial model and the additional training dataset. Moreover, due to $\tilde{\tilde{w}}_j(V_k)$ weighted by the maximum-likelihood estimated method, the results embed the information of the population and lifetime information. Note that, the parameter *Maxiter1* of the iterative time in the symmetrical Kaczmarz method works like the regular parameter.

Main steps for output parameters tuning

The output parameters tuning process proposed in this chapter can be summarized hereafter.

---

1) Based on the maximum likelihood estimation method, forming the linear system problem by formula (4.29).

2) Set $\beta^{[1]} = [a_1 \quad b_1 \quad a_2 \quad b_2 \quad \cdots \quad a_J \quad b_J]^T$ to let the values of output parameters in the initial T-S FIS model as the initial value of the symmetrical Kaczmarz method.

3) Solve the linear system problem iteratively to obtain β' by using the symmetrical Kaczmarz method with *Maxiter1* steps.

Thus, one obtains the input parameters a_j' and b_j' of the tuned T-S FIS model.

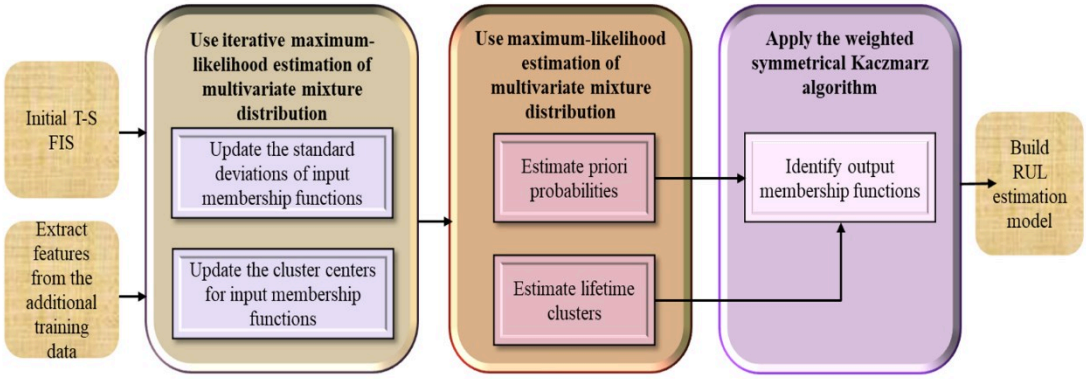


Figure 4.3 The flow diagram of the proposed algorithm for the T-S FIS model tuning.

4.3 Summary

In this chapter, to offset the drawback of a small size sample for model identification we propose a mixture distribution analysis based fuzzy model tuning approach to update the parameters of the T-S FIS model from the initial model by using small-size additional training datasets to obtain a more accurate model. First, the input parameters of the T-S FIS model are updated by using an iterative maximum likelihood estimation of mixture distribution analysis based method which is in this chapter. Then, the output parameters of the T-S FIS model are updated by using another maximum likelihood estimation of mixture distribution analysis based method which is proposed in this chapter by using the symmetrical Kaczmarz method. Finally, the tuned T-S FIS model is formed by using the updated parameters. The number of rules is the same as the initial model.

The main advantage of the approach is that the proposed approach can offset oscillation and non-robust due to small size training data sets. Moreover, the maximum likelihood estimation of the mixture distribution analysis based method proposed in section 4.3.2 embeds information about the population and lifetime of bearings into the output parameters of the tuned model.

Currently, considering the range of the rules influence in the data space, the optimization method of the rules number will be investigated to obtain a more accurate model for estimating the bearings RUL.

Chapter 5 - Case study

The effectiveness of the indicators and T-S FIS identification methods proposed in this work for bearings RUL estimation is assessed using the IEEE PHM 2012 datasets from the PRONOSTIA testbed [226].

In the first subsection, the benchmark dataset is introduced. In the second subsection, the numerical results of the proposed indicators are presented. In the third subsection, the numerical results of the T-S FIS identification method proposed in Chapter 3 are discussed. The numerical results of the T-S FIS tuning method proposed in Chapter 4 are discussed in the fourth subsection. The summary of this chapter is in the fifth subsection.

5.1 The benchmark data

The effectiveness of the T-S FIS identification method introduced in this chapter for bearings RUL estimation is assessed using the IEEE PHM 2012 datasets from the PRONOSTIA testbed, which is provided by FEMTO-ST (Franche-Comté Electronique Mécanique Thermique et Optique – Sciences et Technologies, UMR 6174) Institute in Besançon of France [226].

5.1.1 The PRONOSTIA experimental platform

The PRONOSTIA experimental platform is designed specially to test and validate bearings diagnostic and prognostic approaches by AS2M (Automatique et Systèmes Micro-Mécatroniques) department of FEMTO-ST Institute.

PRONOSTIA can provide real experimental data that characterize the degradation of ball bearings along their whole life cycle until their total failure. Compared with other bearing testbeds proposed in the literature, it allows conducting bearings degradations which contains almost all the types of defects in only a few hours without initiating defects on the bearings. This platform is made up of three main parts: a rotating part, a loading part, and a measurement part.

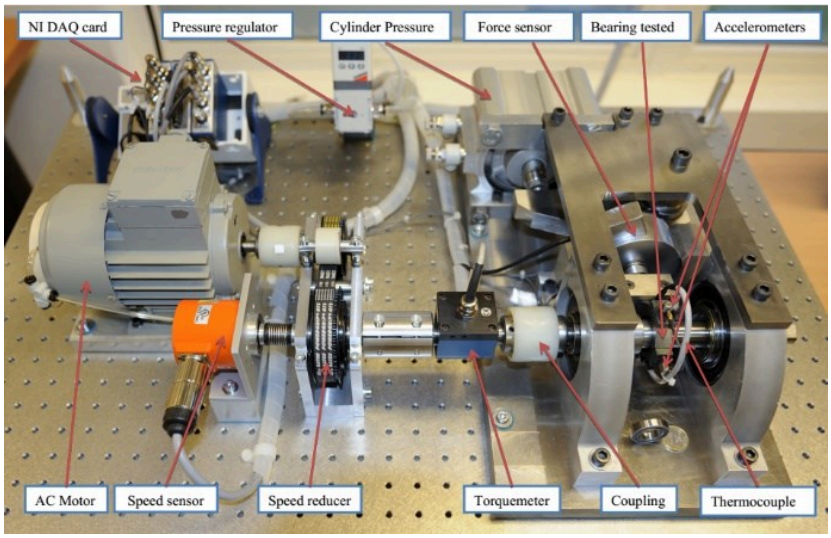


Figure 5.1 Overview of PRONOSTIA [226].

The rotating part includes the asynchronous motor with a gearbox and its two shafts, among which the first one is near to the motor, and the second one is placed at the ride side of the incremental encoder. The motor has a power 250 W and transmits the rotating motion through

a gearbox. Its rated speed can reach 2830 rpm. It can also deliver its rated torque while maintaining the speed of the secondary shaft to less than 2000 rpm. Compliant and rigid shaft couplings are used to create connections for the transmission of the rotating motion of the motor to the shaft support bearing, which leads the bearing through its inner race.

A pneumatic jack, a vertical axis and its lever arm, a force sensor, a clamping ring of the test bearing, a support test bearing shaft, two pillow blocks, and their large oversized bearings are grouped in an aluminum plate to constitute the loading part. The force generated by an actuator in the pneumatic jack is first amplified by a lever arm and is then indirectly applied on the external ring of the test ball bearing through its clamping ring, where a digital electro-pneumatic regulator delivers the supply pressure.

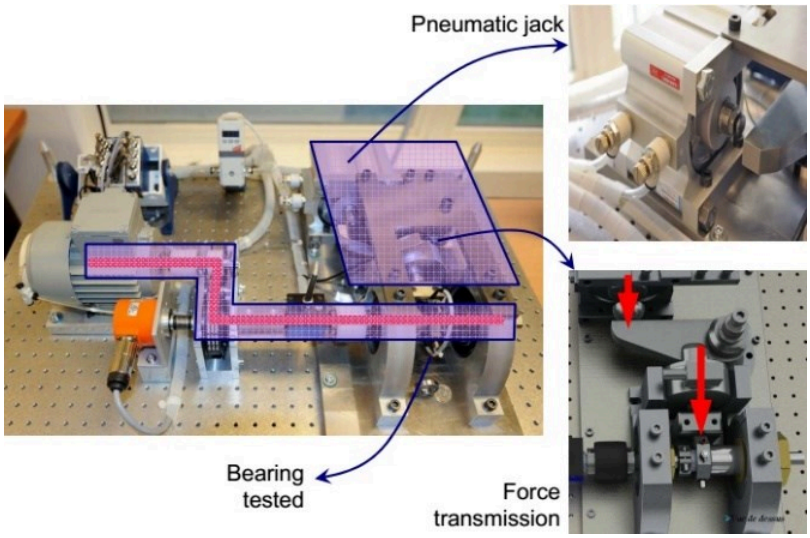


Figure 5.2 The loading part of PRONOSTIA [226].

The measurement part consists of two miniature accelerometers and a temperature sensor. Two miniature accelerometers positioned at 90◦ to each other, i.e., one is placed on the vertical axis, and another is placed on the horizontal axis. The two accelerometers are placed radially on the external race of the bearing. The temperature sensor, a resistance Temperature Detector platinum PT100 (1/3 DIN class) probe, is placed inside a hole close to the external bearing's ring.

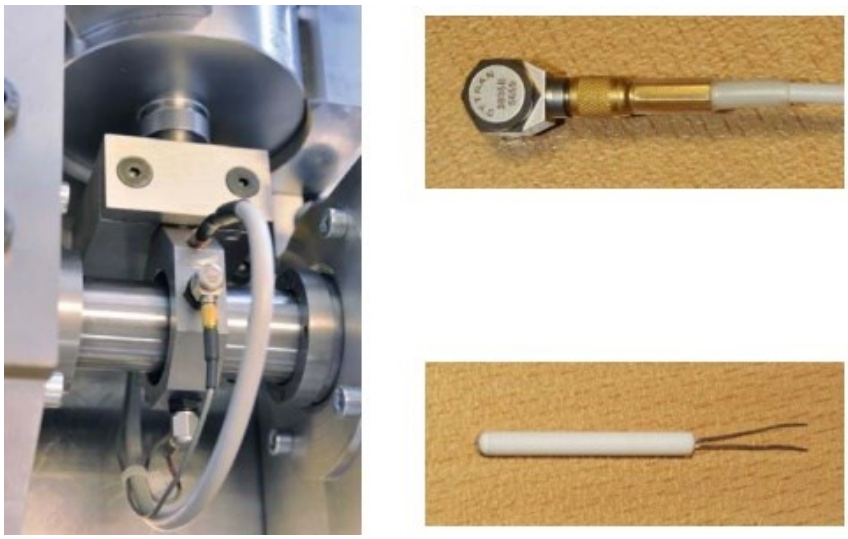


Figure 5.3 The accelerometers and temperature sensor in the measurement part of PRONOSTIA [226].

5.1.2 **Datasets of vibration signals**

The datasets of vibration signals include the data about the same bearings type subjected to three different load conditions. According to [226], two training bearings with the tags (1-1, 1-2) submitted to the load condition 1 and five test bearings (1-3, 1-4, 1-5, 1-6, 1-7) submitted to the same load condition are used in this case study.

The bearings under condition 1 operated at 1800 rpm with 4000 N radial load. Vibration signal data were collected on each of the bearings over an entire run-to-failure period. Two accelerometers have allowed measuring the vibrations in the vertical and horizontal directions. Data were sampled at a 25.6 kHz sampling rate and 0.1s recording duration with 10s sample intervals. Thus, each sample interval is made of 2560 measures. Figure 5.4 displays the sample parameters.

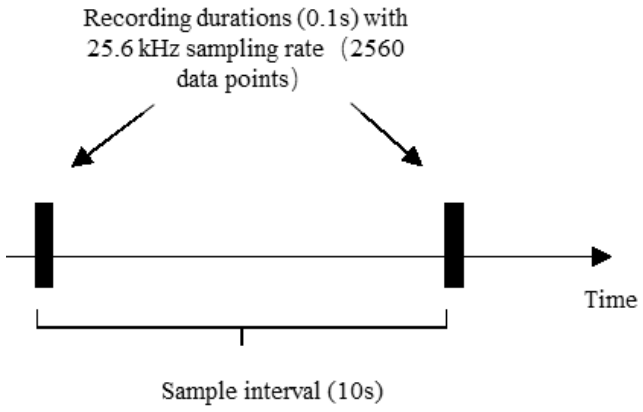
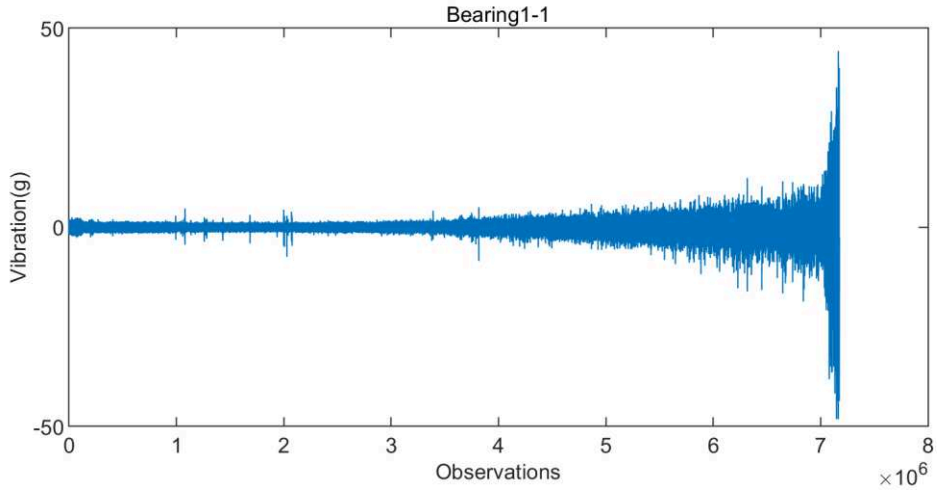
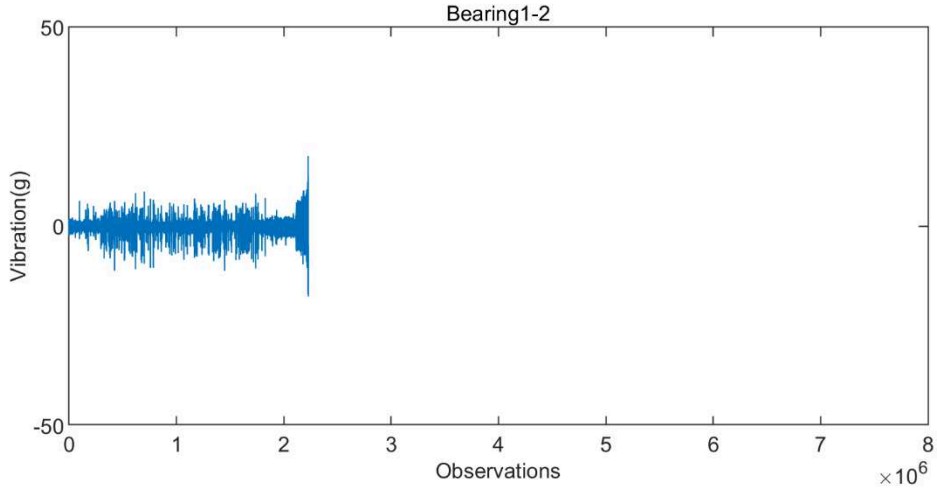


Figure 5.4 Parameters of Sample [226].

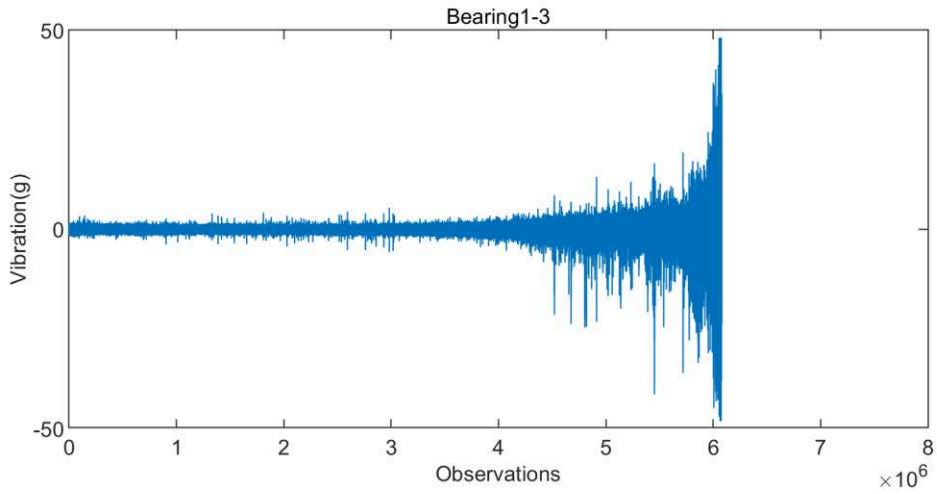
For simplicity, we consider only the vibration signals of the horizontal direction, which are shown in Figure 5.5.



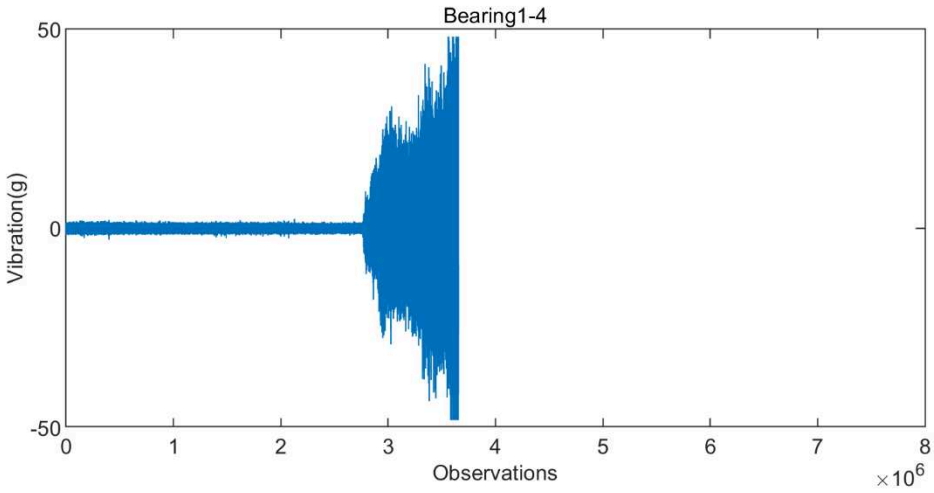
(a)



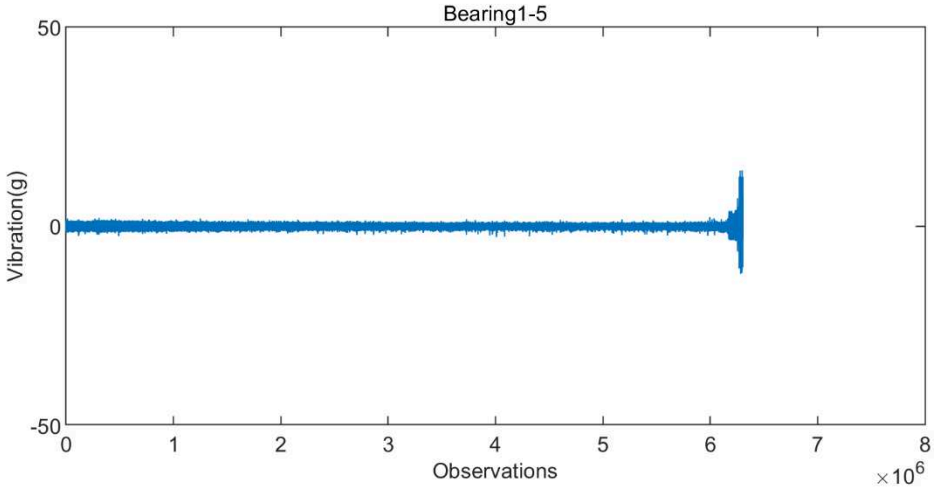
(b)



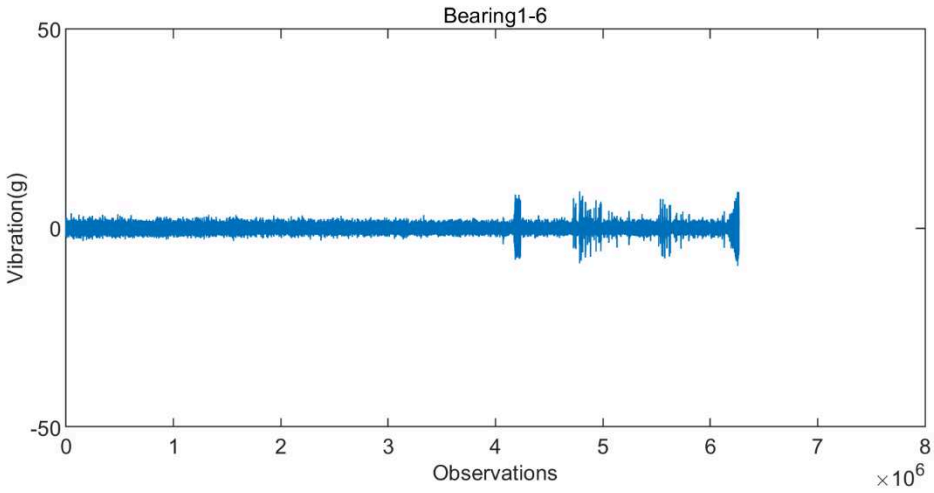
(c)


Bearing1-4
Vibration(g)
50
0
-50
0
1
2
3
4
5
6
7
8
Observations
×10^6

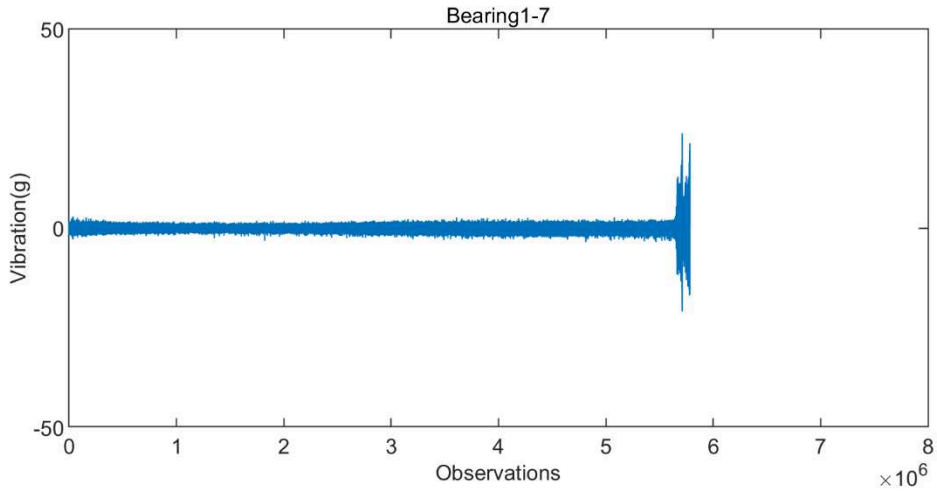
(d)


Bearing1-5
Vibration(g)
50
0
-50
0
1
2
3
4
5
6
7
8
Observations
×10^6

(e)


Bearing1-6
Vibration(g)
50
0
-50
0
1
2
3
4
5
6
7
8
Observations
×10^6

(f)



(g)

Figure 5.5 Bearings vibration signals under condition 1 of the horizontal direction.

The following remarks were highlighted in [226]:

Since the degradation may concern all the components of the bearings at the same time, the theoretical models based on frequency signatures to detect bearings' faults in the inner and outer races, and the cage faults do not work. Existing reliability laws for bearings' life duration, such as the L10, do not give the same results as those obtained by the experiments.

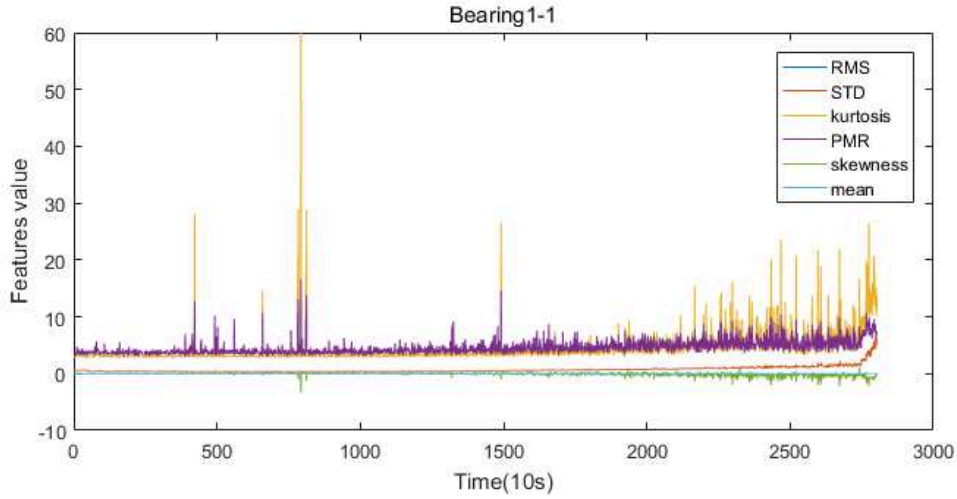
There is no sample information about failure modes and fixed failure thresholds.

5.2 The vibration signals based common features

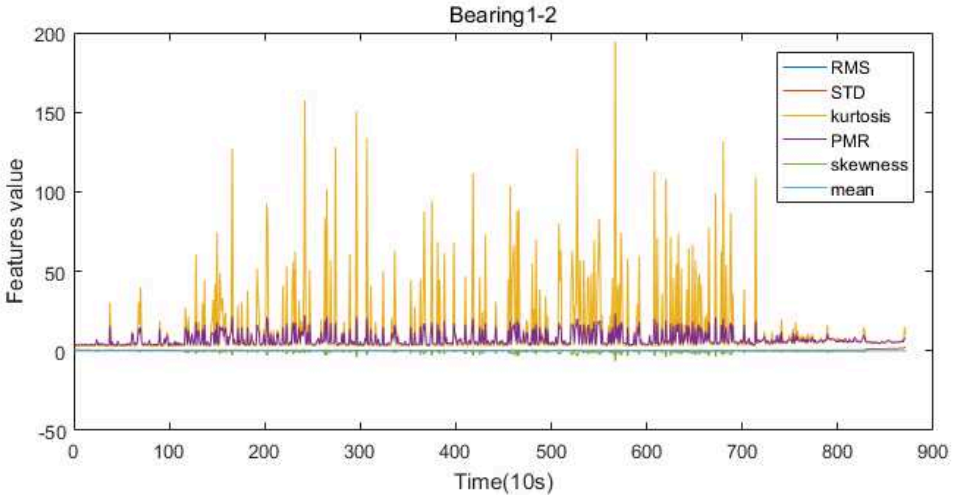
As stated in chapter 2, the common statistical time-domain features we use in this research work are RMS, mean, STD, PMR, kurtosis, and skewness. The RMS, derived from the vibration signals, provides additional information on the energy quantity of the signal. It is an important feature used to estimate the bearing's degradation [19]. It usually increases gradually as the fault develops. The definition of RMS is provided by the formula (1.1). The mean quantifies the mean amplitude of signals, defined in formula (1.2). Note that, for raw vibration signals collected from bearings degradation, the mean remains close to zero but is not zero [20]. This is due to the inherently dynamic characteristic and structural defects of bearings, environmental noises, and irregular shifts of mechanical structures related to bearing degradation. Vibration signals in rolling bearings are non-stationary in nature[21][22]. The STD measures the dispersion of the signals around the mean value. Its definition is provided by the formula (1.3). The PMR quantifies the pluses intensity of the signal, and it is by the formula (1.4). A large PMR value corresponds to a vibration signal with a few large spiky peaks. Similarly, a low

PMR value is relevant to a flattish vibration signal [23]. The mean and standard deviation values of the distribution are the first two parameters of the sample distributions. The third parameter of the distribution is skewness and the fourth is the kurtosis. The kurtosis reveals the degree of deviation of the signal distribution from the average and the skewness indicates the signal distribution asymmetry, their definitions are respectively in formula (1.5) and (1.6).

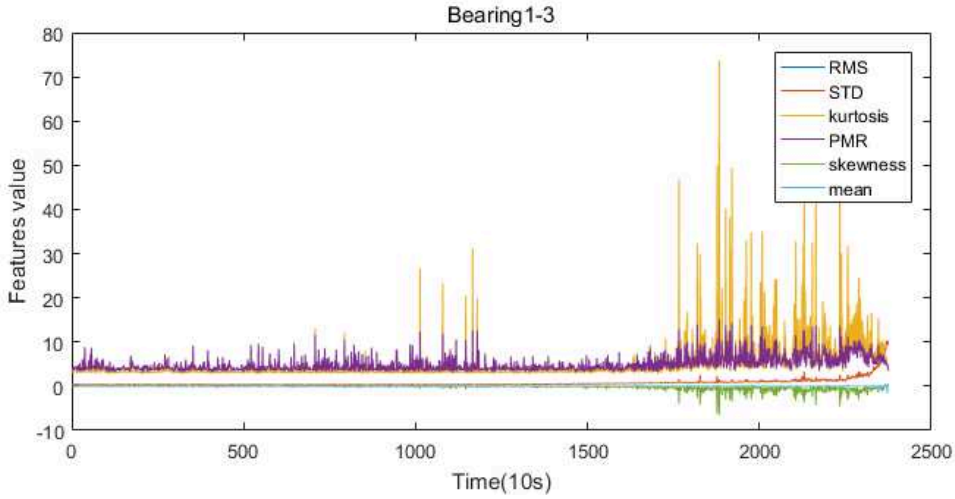
Since the six common statistical time-domain features reveal different characteristics of the vibration signals, we extract indicators based on them. The time series of the six common statistical time-domain features are shown in Figure 5.6.



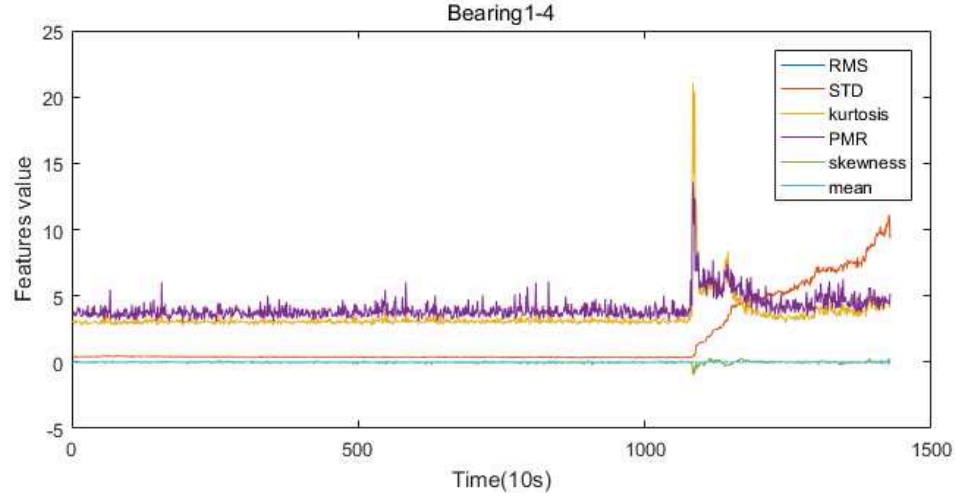
(a)



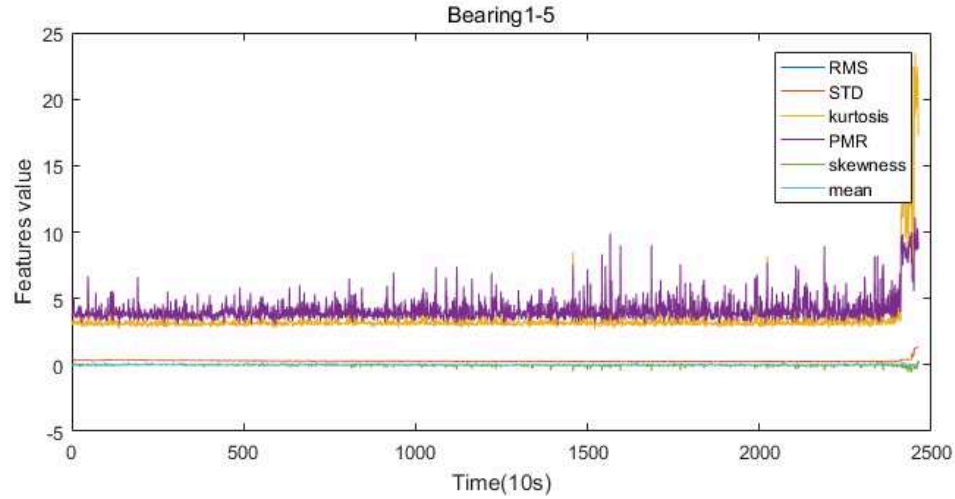
(b)


Bearing1-3
Features value
Time(10s)
RMS
STD
kurtosis
PMR
skewness
mean

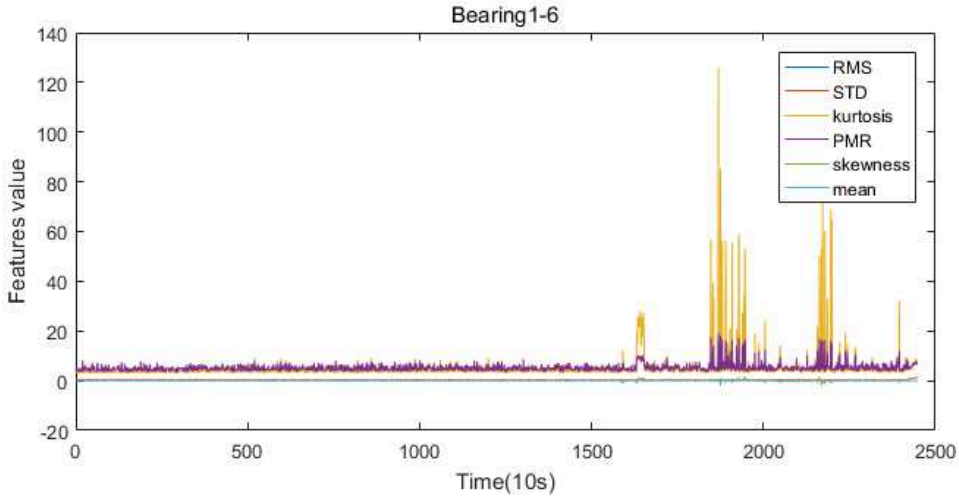
(c)


Bearing1-4
Features value
Time(10s)
RMS
STD
kurtosis
PMR
skewness
mean

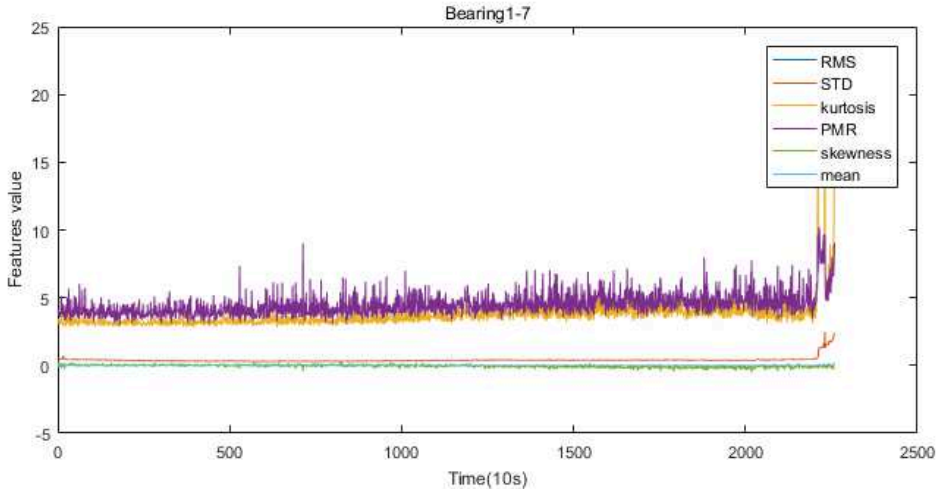
(d)


Bearing1-5
Features value
Time(10s)
RMS
STD
kurtosis
PMR
skewness
mean

(e)



(f)



(g)

Figure 5.6 Six selected common features of the entire lifecycle of bearings.

5.3 Indicators

We chose the data collected from bearings 1-1 and 1-2 to build the training set.

Calculating the indicators requires a certain amount of historical data for the segmentation. Therefore, we extract the indicators starting with an initial sample interval k_{start}, which is equal to 5% of the average number of the sample intervals of the training bearings 1-1 and 1-2 lifetime. Before k_{start}, the health status of the bearings is considered to be the highest.

After processing the time series of the six selected common features over the whole lifecycle of the training bearings1-1 and 1-2 by the PCA method, Figure 5.7 displays the CPV associated with the number of principal components.

It shows that, for the training data, when the number p of principal components is 2, the cumulative percent variance from bearing1-1 and 1-2 are both greater than 95%. Consequently, we choose p=2 for all the bearings when the indicators are being extracted.

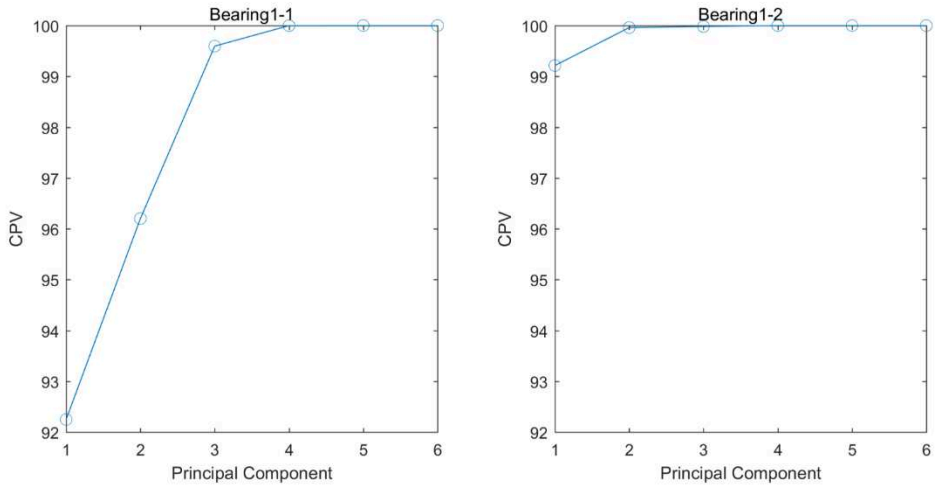
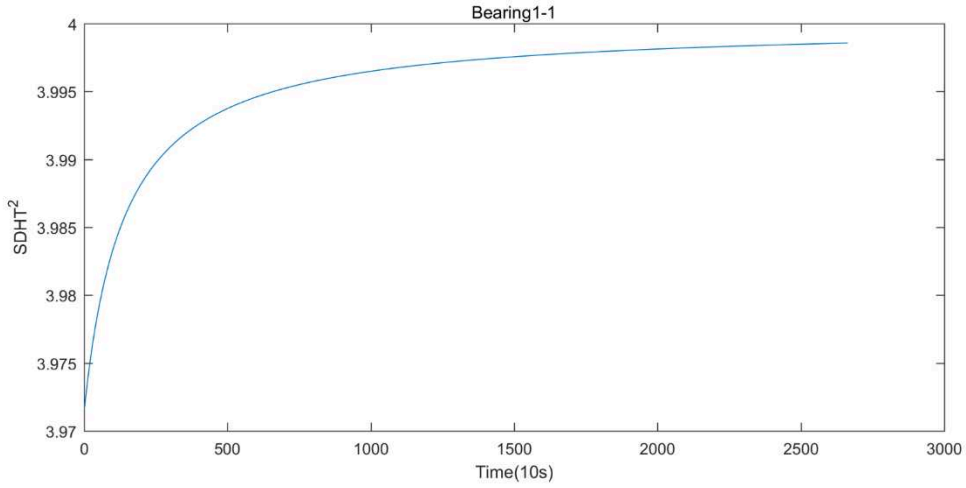


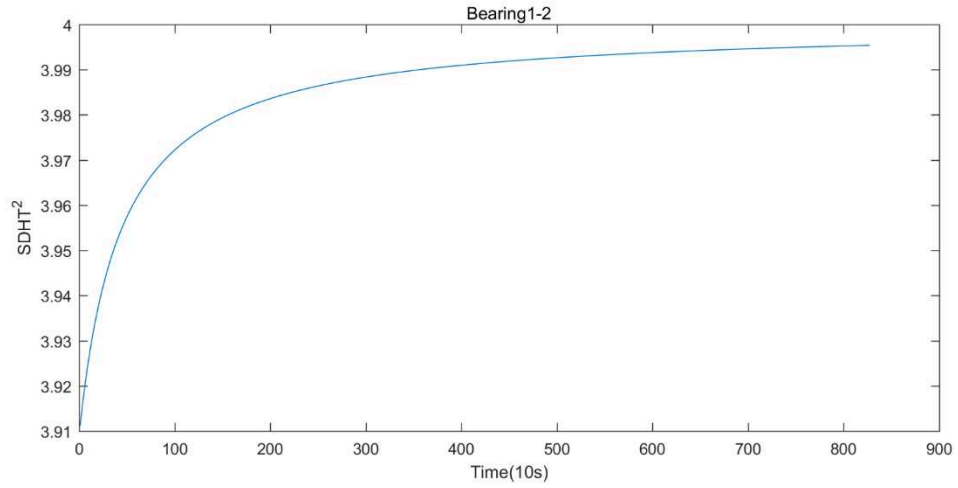
Figure 5.7 Cumulative percent variances.

5.3.1 **Indicators SPE and SDHT^2**

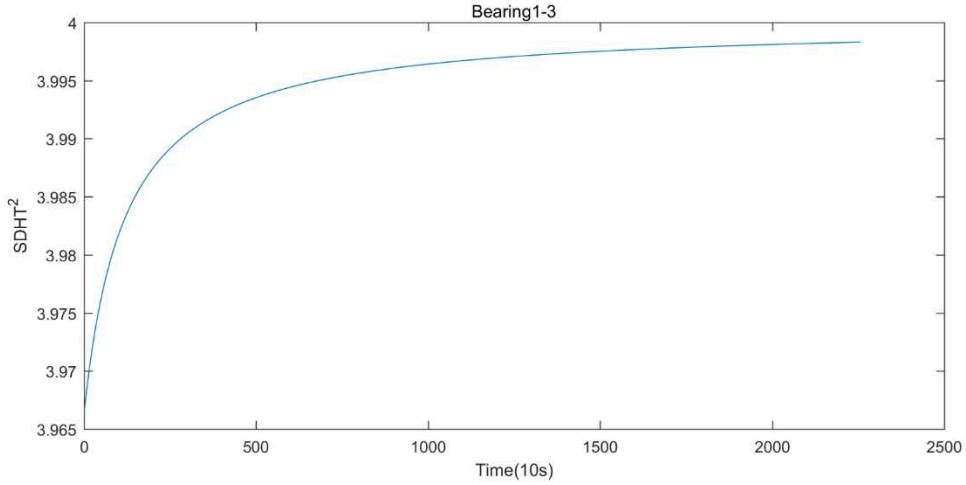
According to the method proposed in chapter 2, the indicator SPE and SDHT^2 extracted from the entire lifecycle of the bearings are respectively plotted in Figure 5.8 and Figure 5.9. The unit of the horizontal axis in Figure 5.8 and Figure 5.9 is 10 seconds. Note that, there is only one value of the indicator corresponding to each 10 seconds.



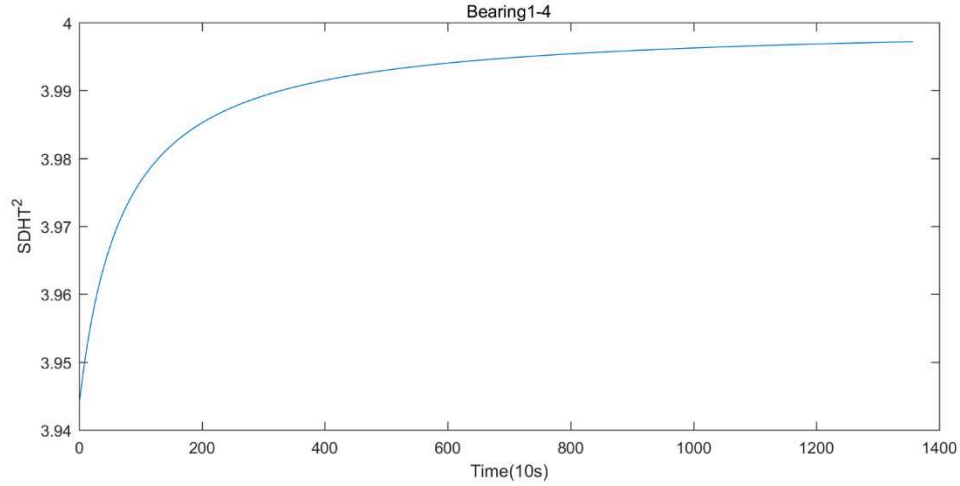
(a)


Bearing1-2
SDHT^2
Time(10s)
4
3.99
3.98
3.97
3.96
3.95
3.94
3.93
3.92
3.91
0
100
200
300
400
500
600
700
800
900

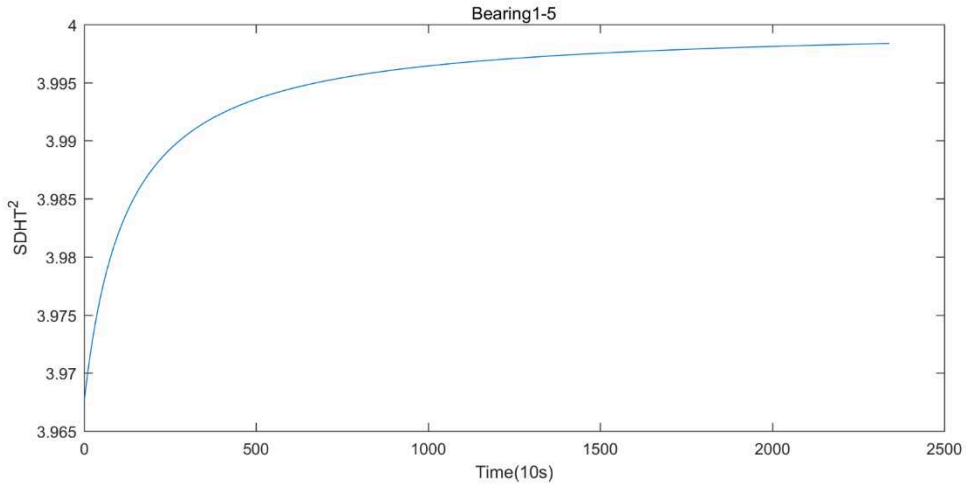
(b)


Bearing1-3
SDHT^2
Time(10s)
4
3.995
3.99
3.985
3.98
3.975
3.97
3.965
0
500
1000
1500
2000
2500

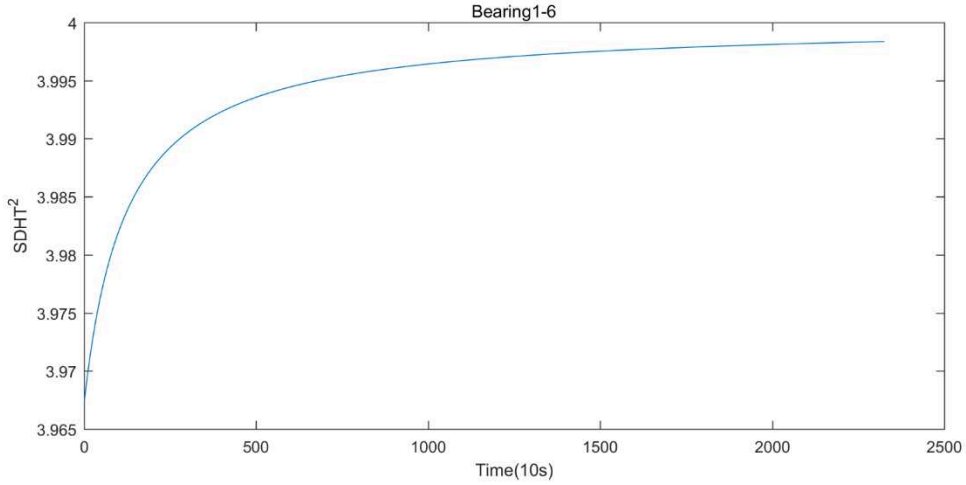
(c)


Bearing1-4
SDHT^2
Time(10s)
4
3.99
3.98
3.97
3.96
3.95
3.94
0
200
400
600
800
1000
1200
1400

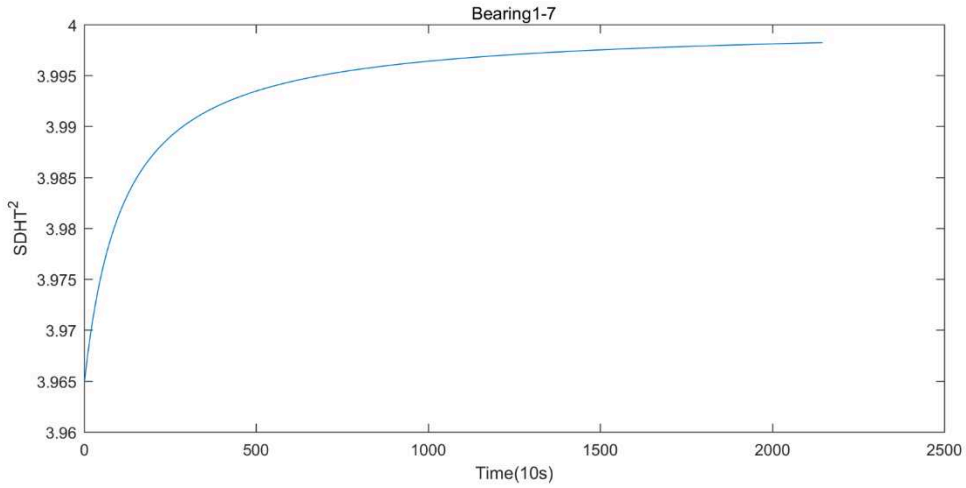
(d)



(e)

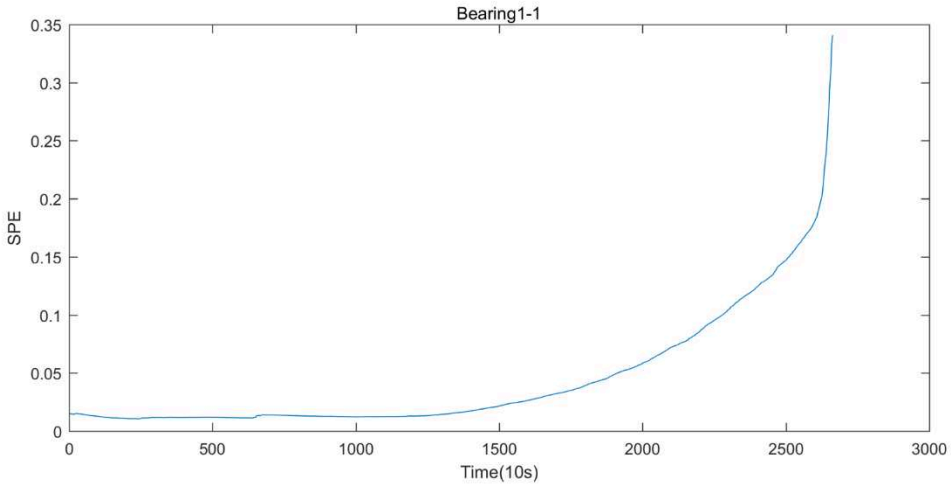


(f)

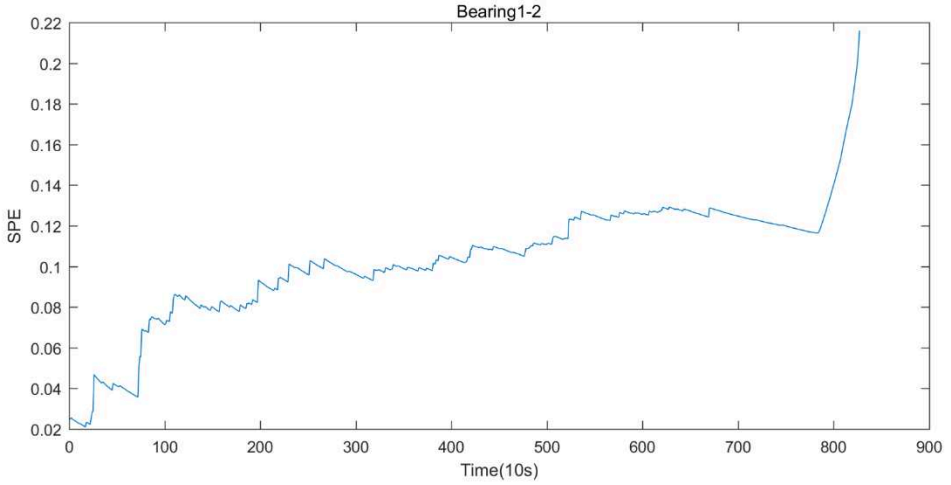


(g)

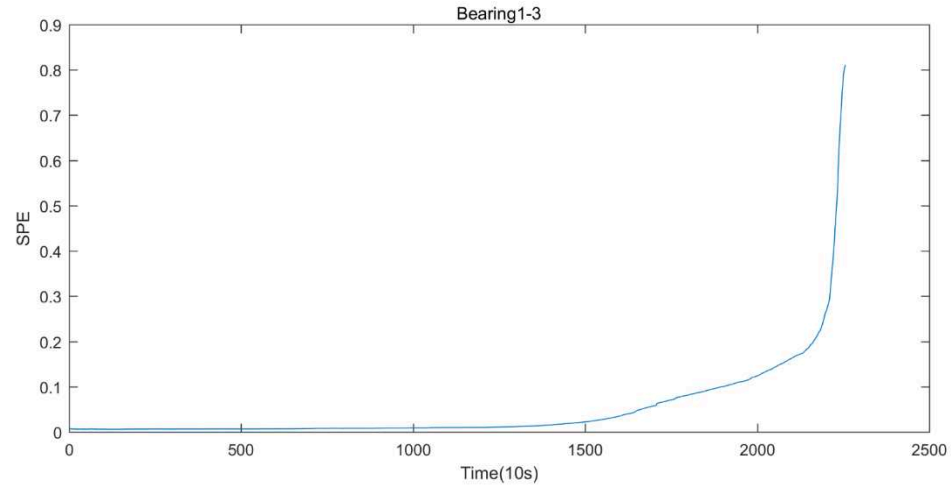
Figure 5.8 $SDHT^2$ indicator of the entire lifecycle of bearings.


Bearing1-1
SPE
Time(10s)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
500
1000
1500
2000
2500
3000

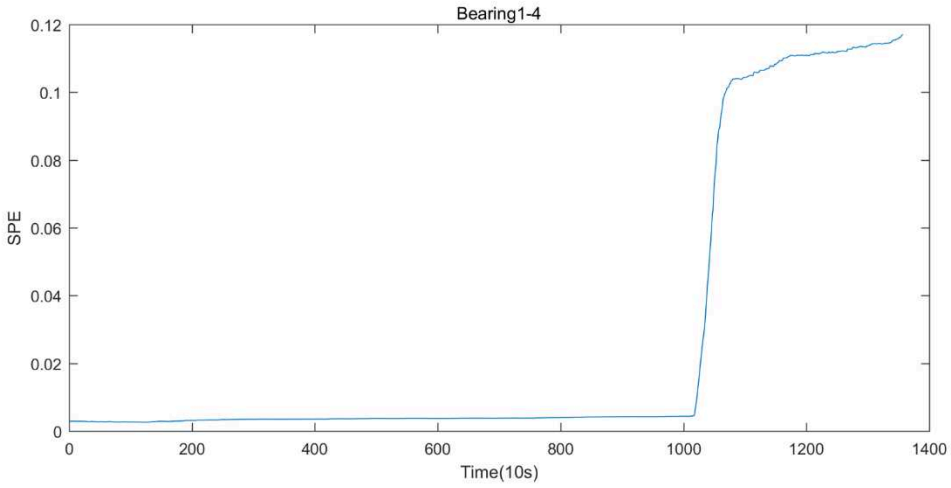
(a)


Bearing1-2
SPE
Time(10s)
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
100
200
300
400
500
600
700
800
900

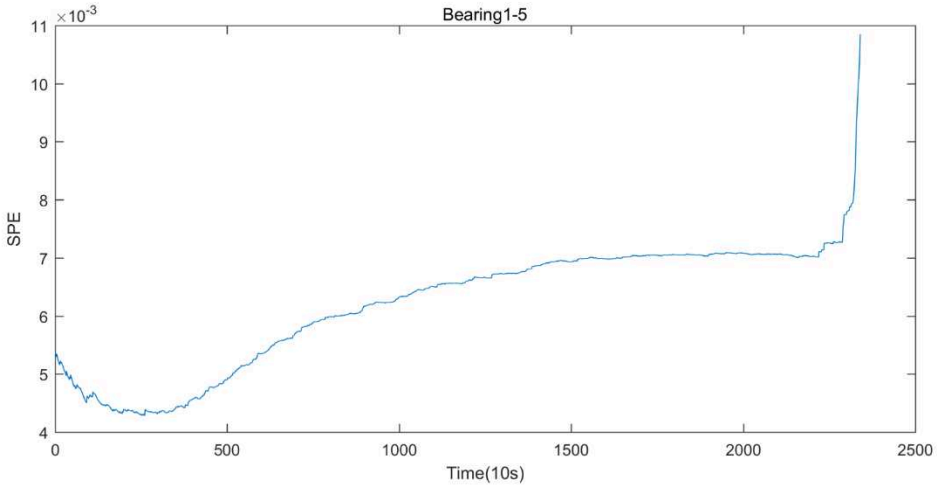
(b)


Bearing1-3
SPE
Time(10s)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
500
1000
1500
2000
2500

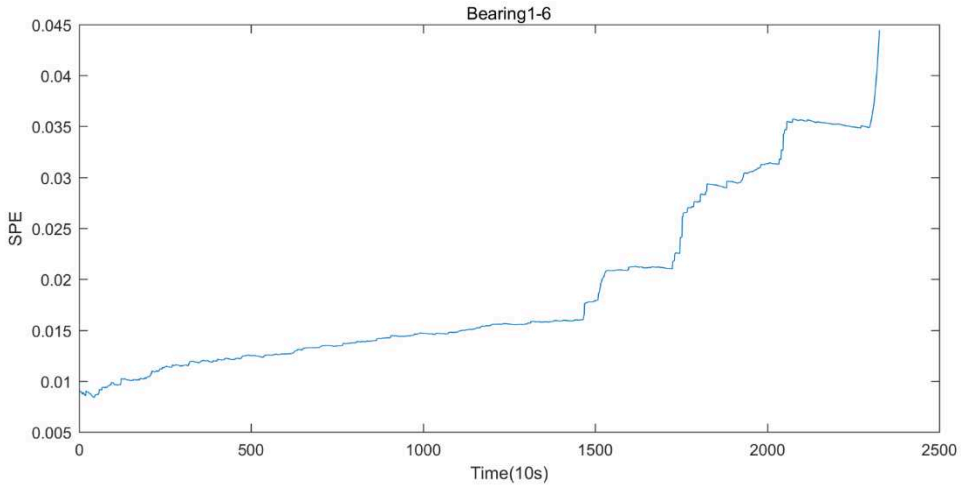
(c)


Bearing1-4
SPE
Time(10s)
0.12
0.1
0.08
0.06
0.04
0.02
0
0
200
400
600
800
1000
1200
1400

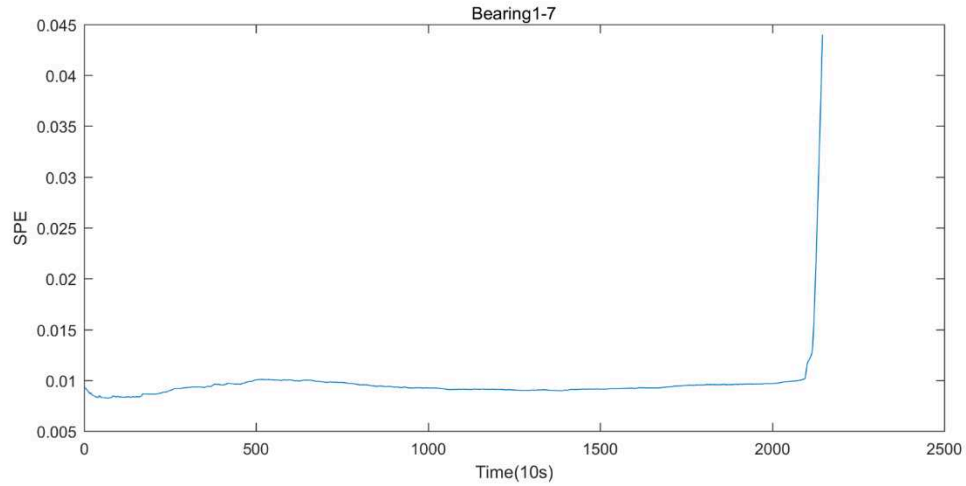
(d)


Bearing1-5
×10-3
SPE
Time(10s)
11
10
9
8
7
6
5
4
0
500
1000
1500
2000
2500

(e)


Bearing1-6
SPE
Time(10s)
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
500
1000
1500
2000
2500

(f)



(g)

Figure 5.9 SPE indicator of the entire lifecycle of bearings.

5.3.2 **Indicators VSPE and VSDHT^2**

According to equations (2.14) and (2.19), the ASDS of different *u*, *z, Q, B* value from the two training bearings 1-1 and 1-2 are displayed in Table 5.1 and Table 5.2. So, in this work, we choose the z as 100, Q as 5 for VSDHT2, u as 200, B as 10 for VSPE. Thus, in this case, the VSDHT2 indicator has 5 dimensions, and the VSPE indicator has 10 dimensions.

Table 5.1 ASDS Value for VSDHT^2

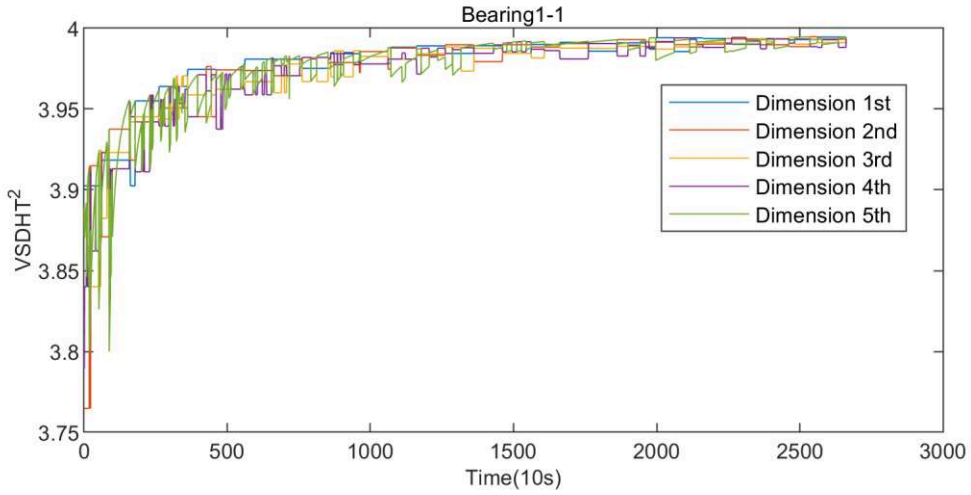
z~Q	**ASDS**
200~20	0.0330
200~10	0.0264
200~5	0.0152
100~20	0.0280
100~10	0.0411
100~5	0.0738
50~20	0.0186
50~10	0.0217
50~5	0.0221

Table 5.2 ASDS Value for VSPE

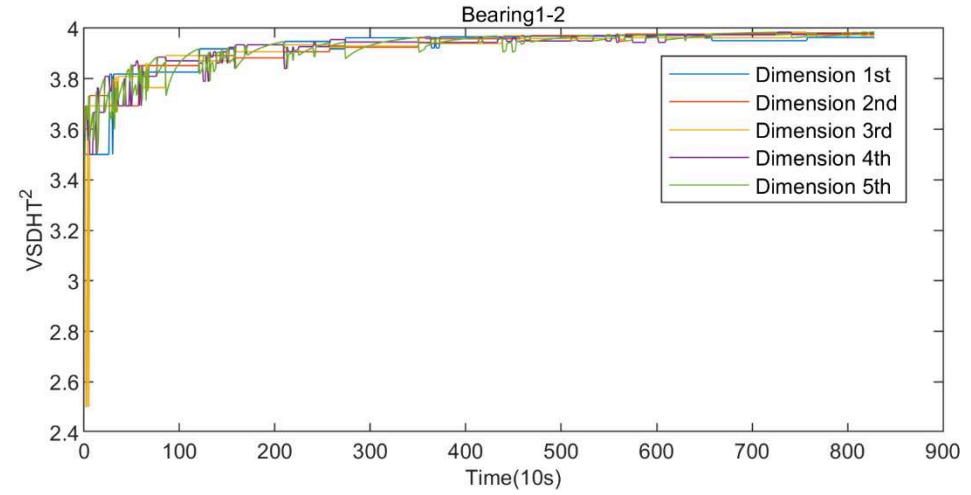
u~B	**ASDS**
200~20	0.0946
200~10	0.1575
200~5	0.1010
100~20	0.0825
100~10	0.1304
100~5	0.1091
50~20	0.0802
50~10	0.0824
50~5	0.1200

According to the method proposed in chapter 2, the indicators VSPE and VSDHT^2 extracted from the entire lifecycle of the bearings are respectively plotted on Figure 5.10 and Figure 5.11. The unit of the horizontal axis in Figure 5.10 and Figure 5.11 is 10 seconds. Note that, there is only a vector of the indicator VSDHT^2, consists of 5 elements, corresponding to each 10 seconds in Figure 5.10. Moreover, there is

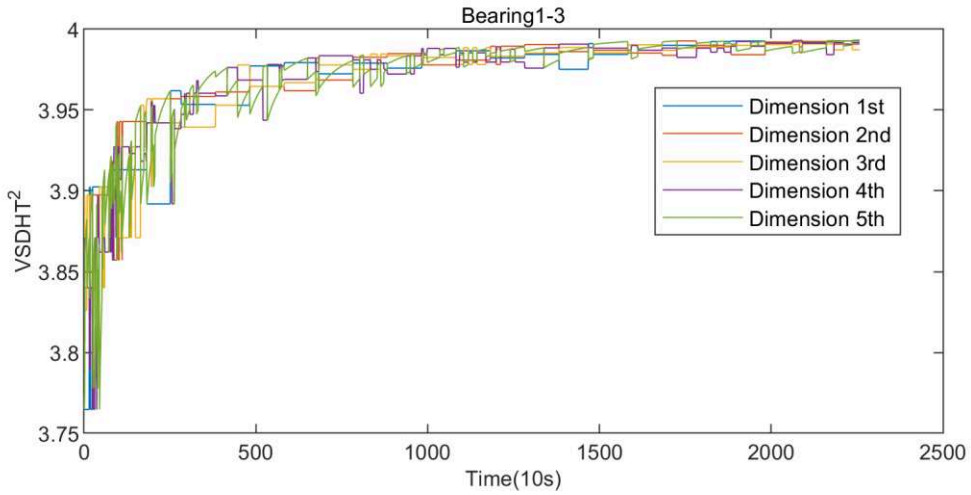
only a vector of the indicator VSPE, consists of 10 elements, corresponding to each 10 seconds in Figure 5.11.



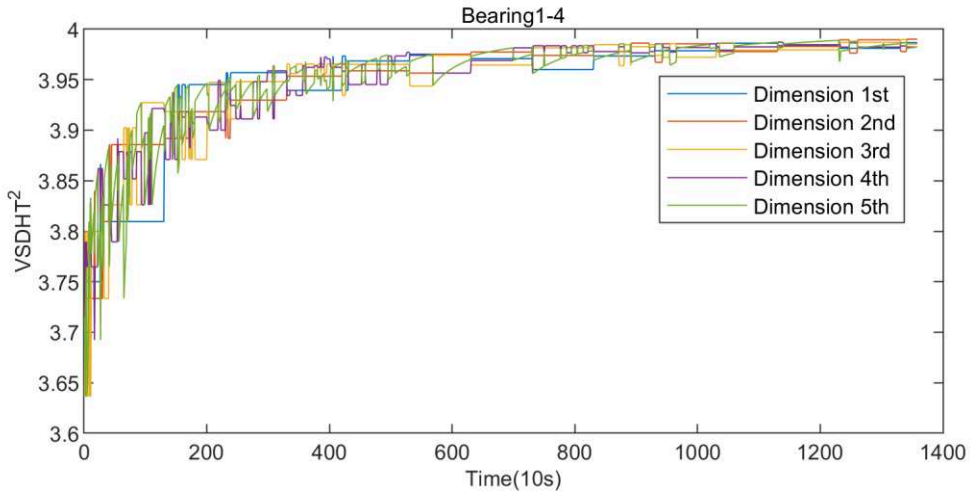
(a)



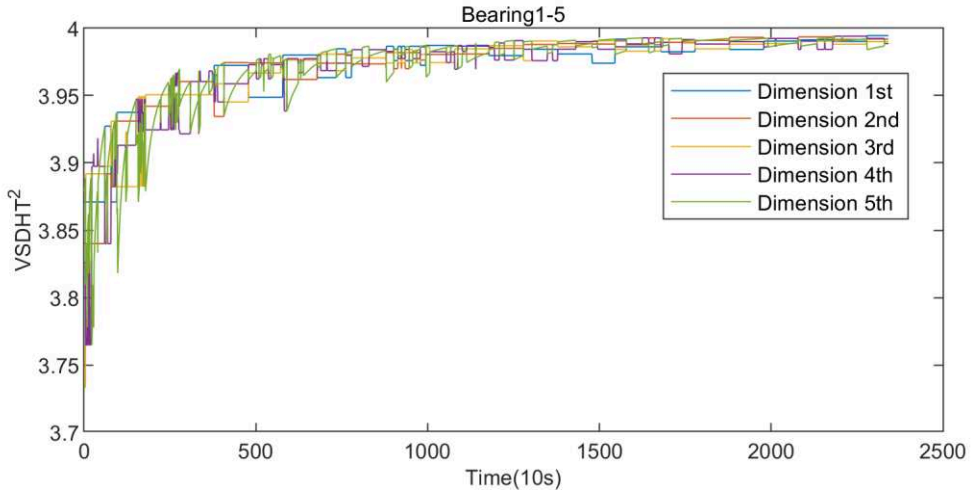
(b)



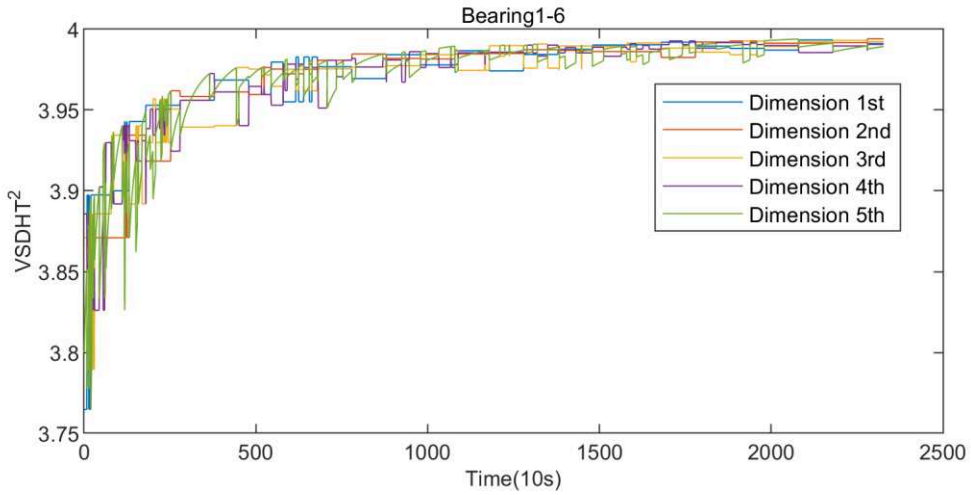
(c)


Bearing1-4
Dimension 1st
Dimension 2nd
Dimension 3rd
Dimension 4th
Dimension 5th
VSDHT2
Time(10s)

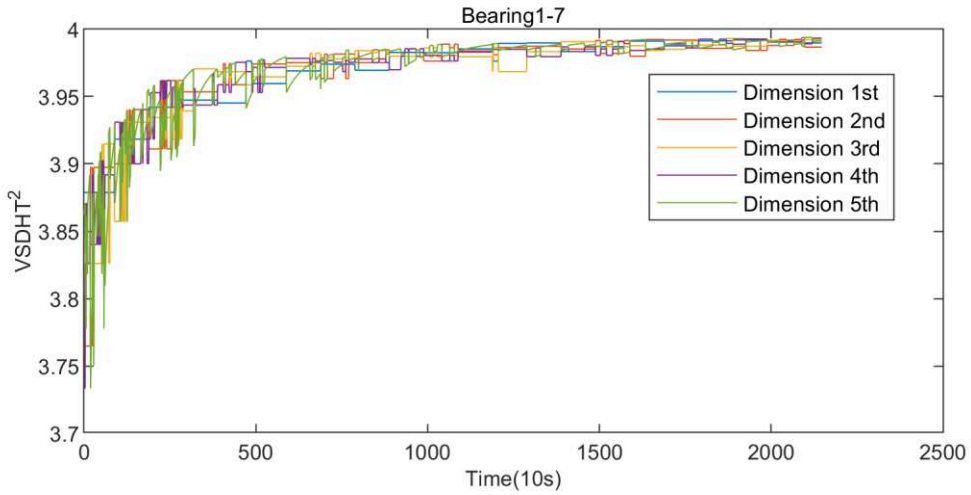
(d)


Bearing1-5
Dimension 1st
Dimension 2nd
Dimension 3rd
Dimension 4th
Dimension 5th
VSDHT2
Time(10s)

(e)

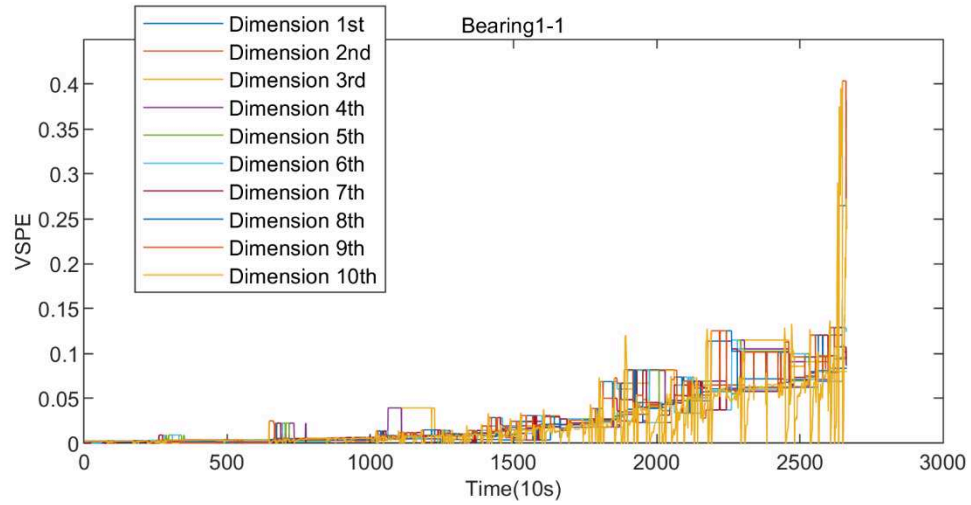

Bearing1-6
Dimension 1st
Dimension 2nd
Dimension 3rd
Dimension 4th
Dimension 5th
VSDHT2
Time(10s)

(f)

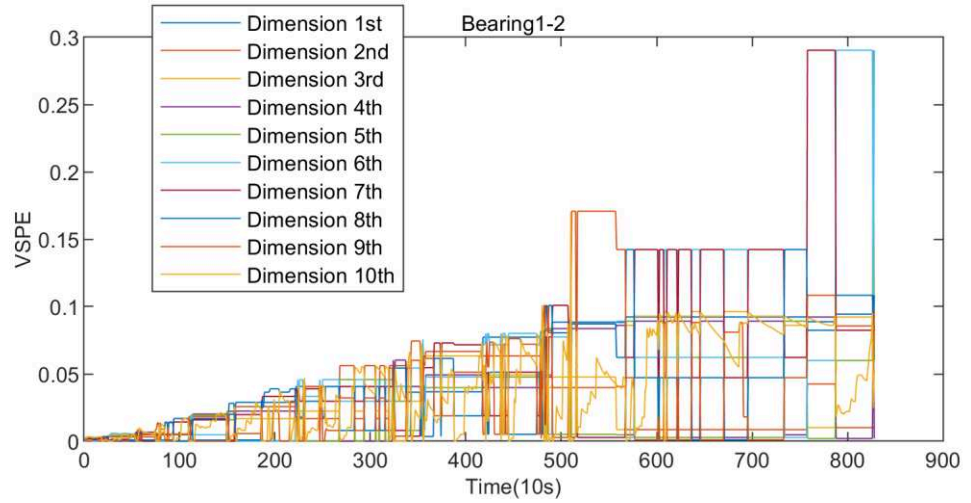


(g)

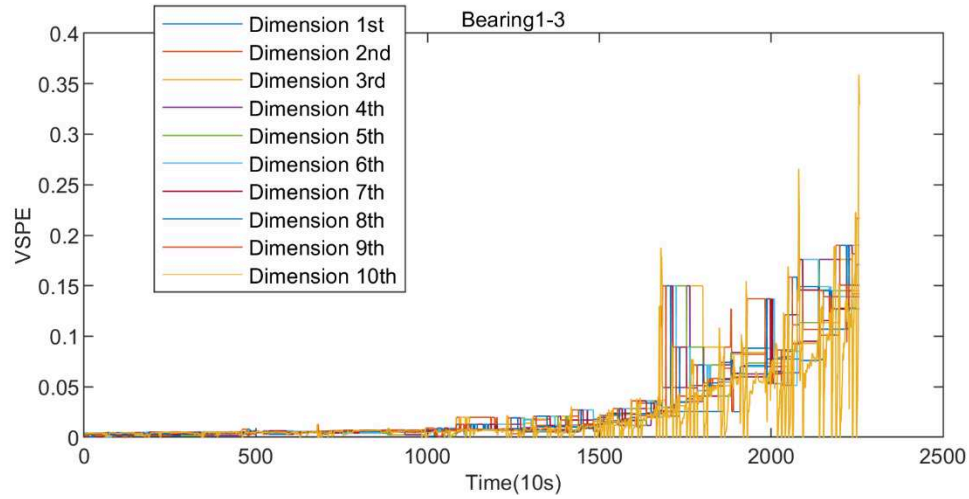
Figure 5.10 VSDHT^2 indicator with 5 dimensions of the entire lifecycle of the bearings.



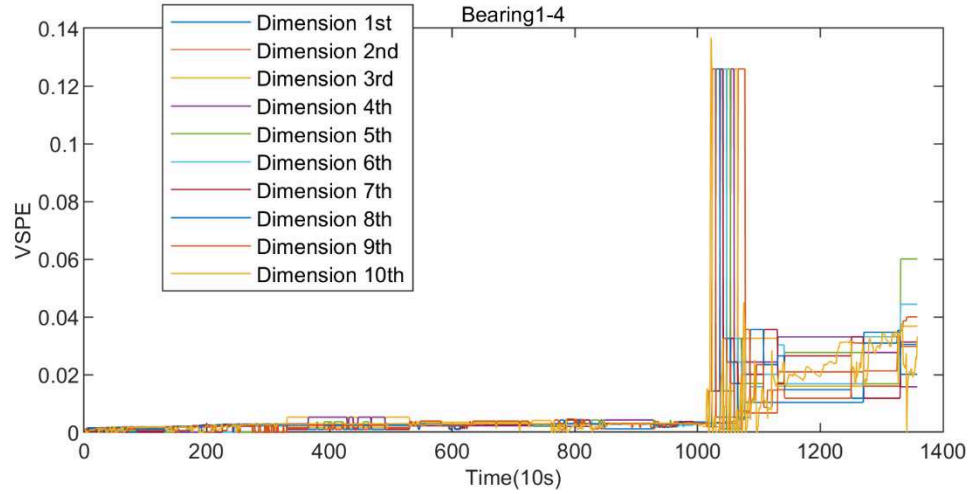
(a)



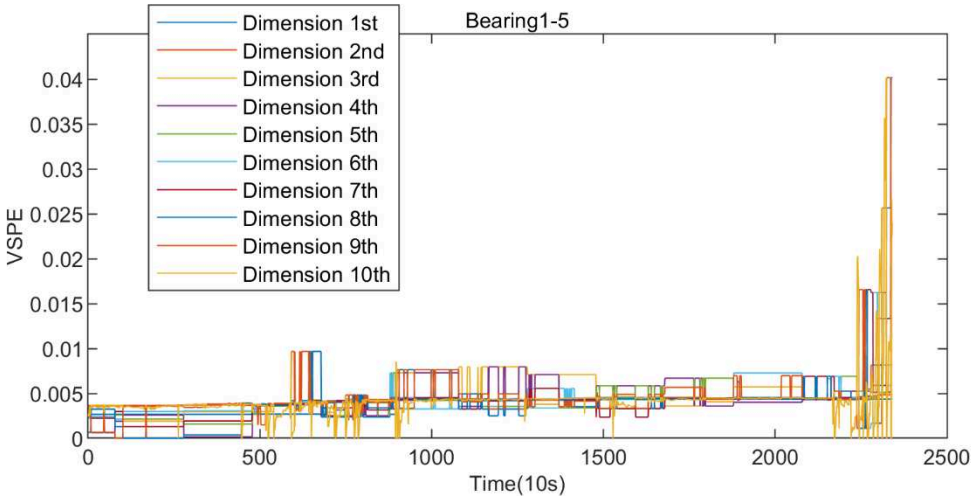
(b)


Bearing1-3
Dimension 1st
Dimension 2nd
Dimension 3rd
Dimension 4th
Dimension 5th
Dimension 6th
Dimension 7th
Dimension 8th
Dimension 9th
Dimension 10th
VSPE
Time(10s)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
500
1000
1500
2000
2500

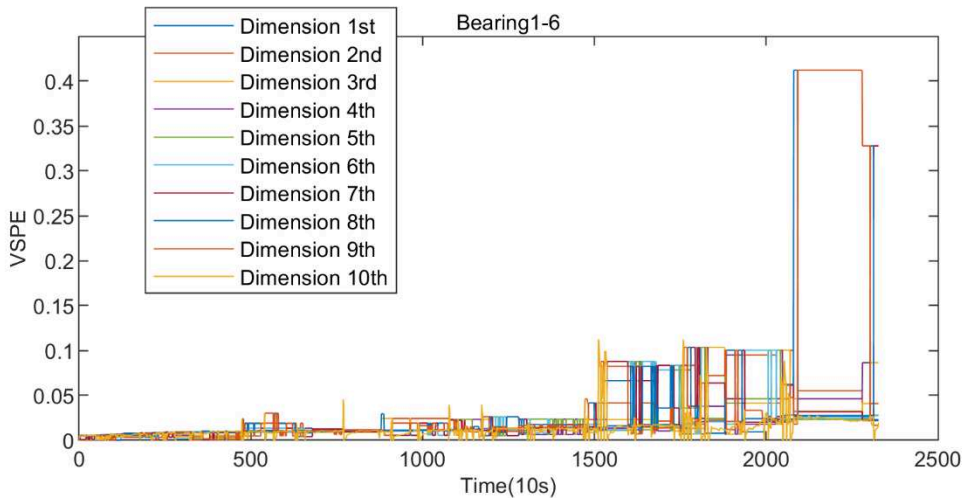
(c)


Bearing1-4
Dimension 1st
Dimension 2nd
Dimension 3rd
Dimension 4th
Dimension 5th
Dimension 6th
Dimension 7th
Dimension 8th
Dimension 9th
Dimension 10th
VSPE
Time(10s)
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
200
400
600
800
1000
1200
1400

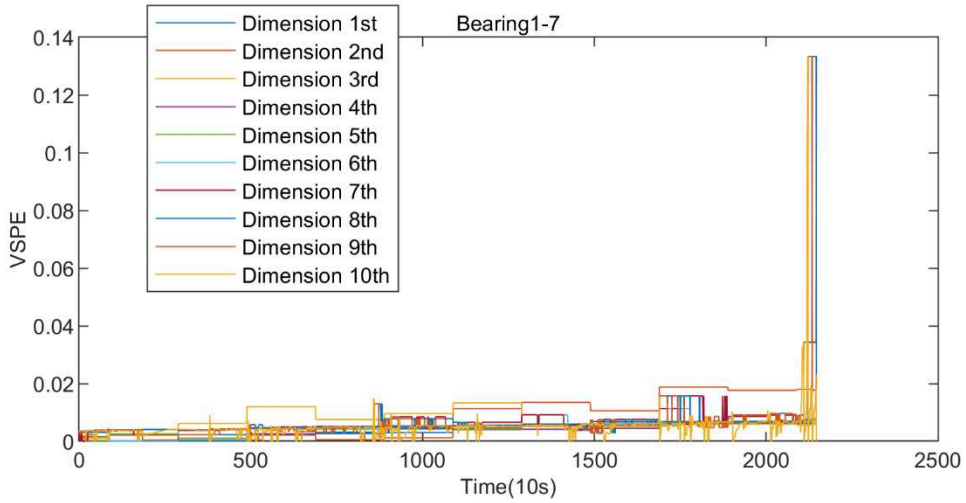
(d)


Bearing1-5
Dimension 1st
Dimension 2nd
Dimension 3rd
Dimension 4th
Dimension 5th
Dimension 6th
Dimension 7th
Dimension 8th
Dimension 9th
Dimension 10th
VSPE
Time(10s)
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
500
1000
1500
2000
2500

(e)



(f)



(g)

Figure 5.11 VSPE indicator with 10 dimensions of the entire lifecycle of the bearings.

According to the results obtained, the new indicators we proposed are all monotonic and continuous sensitivity during the entire life cycle of bearings.

As stated in chapter 2, the new indicators we proposed, can uncover the underlying correlation structure of the multivariate time series of standard degradation features by using the discarded space information to track the degradation evolution of bearings. Moreover, they all highlight the historical degradation information rather than only the current condition. Especially, the indicators VSDHT2 and VSPE, using piecewise linear representation, highlight the historical degradation information, considering the historical process has multiple degradation stages.

5.4 Fuzzy model identification

We use the relative root mean square error (RRMSE) (5.1) and the average of RRMSE (ARRMSE) (5.2) to evaluate the models' performance.

$$RRMSE_o = \sqrt{\frac{1}{K^o}\sum_{k^o=1}^{K^o}\left(\frac{\rho_{k^o} - \hat{\rho}_{k^o}}{\rho_{k^o}}\right)^2} \tag{5.1}$$

$$ARRMSE = \frac{1}{O}\sum_{o=1}^{O} RRMSE_o \tag{5.2}$$

where o=1,2,...O is the test bearings label, k^o=1,2, ...K^o is the sample interval label of o^{th} test bearing.

In general, there are three main categories of identification methods for T-S FIS addressed in this work, e.g., methods based respectively on clustering algorithms, neural networks, and metaheuristic [227]. The identification method introduced in [208] is based on the subtractive clustering and least square estimation (FSC-LSE), which is a representative of the clustering-based identification methods category. The method introduced in [210] is based on the least square estimation and backpropagation gradient descent (BPGD-LSE), which is a representative of the neural networks based identification methods category. Among the metaheuristic algorithms, the genetic algorithm (GA) [228] is the famous population generation based bioinspired metaheuristic algorithm, and the simulated annealing (SA) [229] is one typical single solution based metaheuristic algorithm. For convenience, in the sequel, we shall denote A the FSC-LSE based method, B the BPGD-LSE based method, C the GA based method, and D the SA based method. Also, we shall denote E the method proposed in Chapter 3, based on subtractive clustering with maximum likelihood and weighted least square estimation (MLWLSE).

In this section, the proposed method E is benchmarked with four existing methods for T-S FIS identification, e.g., method A, B, C, and D.

The setting for the subtractive clustering in method A is performed as mentioned in [208]. For method B, the training epoch number is set to 30 steps. The initial step size, step size decrease rate, and step size increase rate for method B are set as suggested in [210]. For the setting of the GA, the population size is set to 200, the mutation function is an adaptive feasible function, crossover fraction is 0.8, and the maximum number of generations is 30. For the SA main parameters setting, the initial temperature is set to100, the maximum number of iterations is set to 30, the reannealing interval is set to 100.

Note that:

1) The identification results of method A is used as initial FIS for the method B, C, and D. The least-square estimation is a non-parametric method, and the backpropagation gradient descent method in method B does not affect the rules. In methods C and D presented, the metaheuristic algorithms only optimize the FIS with parameters of the input and output membership functions but not with the rules. Hence, all models identified by different methods have a consistent structure. Moreover, for methods B, C, and D, the maximum numbers of epochs/ iterations/generations are all set to 30, so to benchmark the results quickly.

2) The models identified by three methods have the same type of input memberships functions (gaussian type), the same number rules, and the same value of $r_j = 1$.

3) All operations are implemented with software MATLAB 2016a, using a computer Dell OptiPlex 7040 equipped with Intel Core i5-6600 CPU and 8GB RAM, Windows 10 Professional OS.

4) The results in Table 5.3 are obtained by using the identified T-S FIS models, with the RMS as the model input.

5) In addition to RMS: The spectral entropy (SE) is a frequency domain feature for bearings degradation assessment [230]. SE can provide a measure of vibration signals' spectral power distribution. The approximate entropy (AE) [231], the largest Lyapunov exponent (LLE) [232], and the correlation dimension (CD) [233], developed based on the chaos theory, are three measurement features of the phase-space dissimilarity for bearings condition monitoring and prognosis [234]. Each value of each feature is obtained from the vibration signal in one sample interval. By processing the vibration signal in k^{th} sample interval, we shall denote SE_k the SE value, AE_k the AE value, LLE_k the LLE value, CD_k the CD value. The results in Table 5.4 are obtained by using the identified T-S FIS models, with the combination of five features above as the model input.

6) The results in Table 5.5 are obtained by using the identified T-S FIS models by using the combination of VSPE and VSDHT^2 as the model input of models. The combination of VSPE and VSDHT^2 means to serial concatenate the two vectors together to form a new vector.

7) Bearings 1-1 and 1-2 are used as the training set, bearings 1-3,1-4,1-5,1-6,1-7 are used as test set.

Table 5.3 The results with input RMS (V_k=RMS_k)

Identification method	RRMSE					ARRMSE	Execution time
	Test bearing 1-3	Test bearing 1-4	Test bearing 1-5	Test bearing 1-6	Test bearing 1-7		
Method A (FSC-LSE)	1.2362	1.5269	1.2433	1.4719	1.3047	**1.3566**	**2.4219s**
Method B (BPGD-LSE)	1.2216	1.4544	1.2391	1.4600	1.2987	**1.3348**	**2.7031s**
Method C (GA)	1.2928	1.4564	1.3046	1.4559	1.3476	**1.3715**	**239.5469s**
Method D (SA)	1.6325	1.7260	1.6302	1.7218	1.6591	**1.6739**	**6.6094s**
Method E (MLWLSE)	0.6979	0.8263	0.8106	0.8556	0.7991	**0.7979**	**2.4844s**

Table 5.4 The results with input RMS (V_k=[RMS_k, SE_k, AE_k, LLE_k, CD_k]^T)

Identification method	RRMSE					ARRMSE	Execution time
	Test bearing 1-3	Test bearing 1-4	Test bearing 1-5	Test bearing 1-6	Test bearing 1-7		
Method A (FSC-LSE)	0.8939	0.5819	0.7837	1.1346	0.7752	**0.8339**	**3.1719s**
Method B (BPGD-LSE)	0.8657	0.9527	0.7137	0.8153	0.5998	**0.7894**	**5.7344s**
Method C (GA)	0.8949	0.5761	0.8876	1.1057	1.0797	**0.9088**	**1043.9656s**
Method D (SA)	1.0634	0.7764	0.9535	1.2657	1.267	**1.0652**	**12.5938s**
Method E (MLWLSE)	0.6436	0.5697	0.6879	1.0032	0.5283	**0.6865**	**3.3906s**

Table 5.5 The results with input combination of VSPE and VSDHT^2 (V_k=[$VSPE_k$; $VSDHT^2_k$])

Identification method	RRMSE					ARRMSE	Execution time
	Test bearing 1-3	Test bearing 1-4	Test bearing 1-5	Test bearing 1-6	Test bearing 1-7		
Method A (FSC-LSE)	0.1585	0.4443	0.3559	0.3331	0.2585	**0.3101**	**4.7969s**
Method B (BPGD-LSE)	0.1250	0.4811	0.3448	0.2997	0.2907	**0.3083**	**12.0469s**
Method C (GA)	0.2964	0.5577	0.4648	0.5441	0.3032	**0.4332**	**1695.4594s**
Method D (SA)	0.4943	1.3974	0.5939	0.5087	0.6667	**0.7322**	**18.1716s**
Method E (MLWLSE)	0.1720	0.4450	0.3557	0.2637	0.2956	**0.3064**	**5.2344s**

Based on the same input, according to the results in Table 5.3 and Table 5.5:

1) Method C and D consume more execution time to yields acceptable results.
2) Method A consumed the least time to identify the T-S FIS model among the five methods.
3) Method B and E consume extra execution time while obtaining models that are more accurate compared with method A. Moreover, method E is computationally effective with more accurate results, relatively to Method B.
4) Method E is the one that identifies the most accurate T-S FIS model among the five methods.
5) Due to the merits of VSPE and $VSDHT^2$ for bearings degradation assessment, the performance of models is improved a lot.

The research work presented in Chapter 3 deals with TS-FIS model identification using a data set collected from a small number of bearings. These numerical results derived on this data set show that the method developed in chapter 3 is promising for TS-FIS model identification for a small amount of historical data.

Additionally, the numerical results reveal that using a set of features, especially the combination of VSPE and $VSDHT^2$, provides better performance than using only one feature. According to these results, the proposed indicators VSPE and $VSDHT^2$ are promising for bearings degradation monitoring.

5.5 Fuzzy model tuning

For comparison, an existed famous method for T-S FIS model identification and tuning method B are also used in this subsection. Let method F denoted the tuning approach for the T-S FIS model proposed in Chapter 4 (Tuning Approach).

In this section, the proposed method F is benchmarked with four existing methods for T-S FIS identification, e.g., methods B, C, and D.

Note that:

1) The training epoch number for Method B is set as 60. The initial step size, step size decrease rate, and step size increase rate for Method B are set according to [210]. In methods C and D presented, the metaheuristic algorithms only optimize the FIS with parameters of the input and output membership functions but not with the rules. Hence, all models identified by different methods have a consistent structure. Moreover, for methods B, C, and D, the maximum numbers of epochs/ iterations/generations are all set to 60, so to benchmark the results quickly.

2) The models under four methods have the same type of input memberships (gaussian type), the same number rules, and the same value of r_j as 1.

3) For method F, the parameter $Maxiter$ is set as 1000, and the parameter *Maxiter1* is set as 10. The relax parameter δ of the symmetric Kaczmarz algorithm is set to 1.

4) The models obtained for Table 5.5 are the initial T-S FIS models to corresponding methods in Table 5.6 due to its accuracy. Hence, the combination of VSPE and $VSDHT^2$ is used as the input of the model.

5) Based on bearings 1-3,1-4,1-5,1-6,1-7, leave one bearing out cross-validation method is used to establish the additional training set and test set for tuning methods.

Table 5.6 Tuned T-S FIS model under different methods by leaving one bearing out cross-validation (V_k=[$VSPE_k$; $VSDHT^2_k$])

Tuning method	RRMSE					ARRMSE	Execution time
	Test bearing 1-3	Test bearing 1-4	Test bearing 1-5	Test bearing 1-6	Test bearing 1-7		
Method B (BPGD-LSE)	0.8816	0.6985	0.2742	0.7247	0.2565	**0.5671**	**147.31s**
Method C (GA)	1.1181	2.5817	0.4626	0.9721	1.1849	**1.2639**	**19229.95s**
Method D (SA)	0.4943	1.3974	0.5939	0.5087	0.6667	**0.7322**	**130.11s**
Method F	0.1891	0.3882	0.3235	0.2818	0.305	**0.2975**	**120.50s**

According to the results in Table 5.6, using additional training sets, the models tuned by method F have better performance. Method F appears to spend less execution time compared to method B. Compared with the results in

Table 5.5, the results of the tuning process by methods B, C, and D in Table 5.6 are worse. In contrast, the results of the tuning process by method F in Table 5.6 is better, according to the values of ARRMSE. These results prove that the parameters tuning process implemented by the methods B, C, and D is affected by the small size of the additional training data set. The proposed method F is promising for tuning T-S FIS models to estimate bearings RUL.

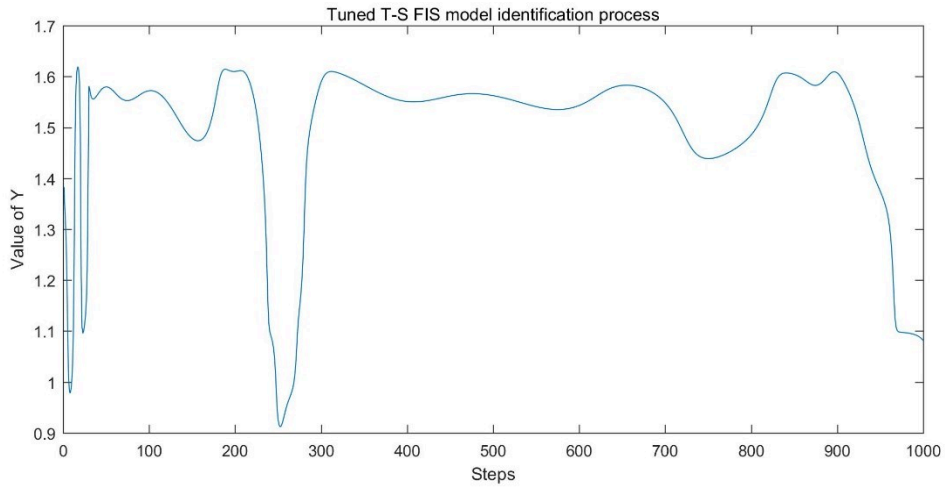


Figure 5.12 The values of Y along the identification process of the tuned T-S FIS model corresponding to leave bearing 1-5 out.

Additionally, the iterative process in Method F has a light calculation load. One can set a relatively big number of steps for parameter $Maxiter$ to find a relatively better local optimum of the value of $c_{j,i}^{'}$. Besides, the number of iterative steps works like the regular parameter. The evolution of clusters' validity index Y defined in equation (4.15) of subsection 4.2, along the identification process of tuned T-S FIS model corresponding to leave bearing 1-5 out is shown in Figure 5.12. Hence, among a relatively big number of steps for the identification process of the tuned T-S FIS model, a relatively better local optimum of the value of $c_{j,i}^{'}$ could be found.

5.6 Summary

The evolutions of bearings degradation process are complex and diverse. The vibration signals collected over the entire run to failure periods of bearings have heavy background noise. Additionally, the processes of bearing degradation are complex. Hence, the process of mining the information embedded in raw vibration signals is necessary.

A valuable characteristic of the new indicator is to be monotonic through bearings' entire lifecycle. This requirement is natural, knowing that bearings are mechanical components and self-repair is improbable during periods of non-utilization.

According to the numerical results in this chapter, one can observe that: the six vibration signals based common features are sensitive and monotonic only during certain stages of the bearings degradation process.

The proposed indicators in this chapter are sensitive and monotonic during the entire run to failure period of bearings. Especially, the proposed indicators VSPE and VSDHT^2 can also

carry both the current and historical information of bearings degradation due to their different dimension associated with different stages of the historical degradation process of bearings. The proposed method of T-S FIS model identification and tuning can achieve a more accurate model while spending relatively less execution time.

General conclusion

This thesis has progressively established an effective approach for identical bearings prognosis in case of small size available sample data from bearings under identical load. This research brings several contributions.

The first contribution consists of working out indicators which can efficiently characterize the degradation of bearings, in this work. We propose methods respectively based on the $SDHT^2$ and SPE with a piecewise linear representation approach. First, we used time-domain common features, extracted from the bearing vibration signal, to describe the bearing degradation roughly. Next, characteristic values are used to divide the whole historical degradation process into homogenous segments by processing the vibration signal based time-domain common features through the piecewise linear representation approach. Then, characteristic values are calculated from all homogenous segments. Finally, a vector consists of these characteristic values, as its entries, is the indicator. When the $SDHT^2$ values are used as the characteristic values, the indicator is called $VSDHT^2$. When the SPE values are used as the characteristic values, the indicator is called VSPE. By a numerical study using benchmark datasets, the indicators data series show a monotonic trend during the bearing degradation and is sensitive to the degradation all along the bearings' lifetime. Furthermore, the new indicators carry not only the information corresponding to the bearings' current state but also the historical evolution of bearings' degradation.

The second contribution consists of a proposed method to identify a simple T-S FIS for bearings RUL estimation, which could mine more useful information associated with population and lifetime of bearings from data to obtain accurate models with a small number of parameters. Each rule in the T-S FIS model is an application that associates a degradation feature input data to one, and only one fuzzy subset of degradation and the corresponding lifetime. The proposed method is effective with small historical data. To that extent, we used the vibration signals collected over the complete run to failure lifecycle of a small number of bearings over an entire period of run-to-failure. The training set was built using features extracted from periodic observations of the vibration signals. The number of rules and the input parameters of each rule were identified using the fuzzy subtractive clustering method. Furthermore, we proposed to use the maximum likelihood method of mixture distribution analysis to calculate the maximum likelihood clusters on the time axis and priori probability of maximum likelihood corresponding

to each degradation stage. Based on this result, we identified the output parameters of each rule using a weighted least square estimation.

The third contribution consists of a proposed tuning approach for tuning the T-S FIS model to obtain a more accurate model for bearings RUL estimation in case of small size additional training data sets derived from the vibration signal over the complete run to failure lifecycle of a small number of identical bearings under identical load. The approach could offset the oscillation and non-robustness of the model tuning process due to the small size of training data sets. First, the input parameters of the T-S FIS model are updated by using an iterative maximum likelihood estimation of mixture distribution analysis based method which is in this chapter. Then, the output parameters of the T-S FIS model are updated by using another maximum likelihood estimation of mixture distribution analysis based method which is proposed in this chapter by using the symmetrical Kaczmarz method. Finally, the tuned T-S FIS model is formed by using the updated parameters.

Our research work still could be expanded its scope of applications. First, some appropriate measurements could be used with the piecewise linear representation method to build degradation indicators to reveal the underlying information among in vibration signals of bearings with an unidentical load.

Second, considering the influence of each rule of the fuzzy model in the data space, rules addition and removal could be conducted along with the parameters tuning process by some appropriate methods to optimize the fuzzy model with additional data samples.

References

[1] K. F. Martin, “A review by discussion of condition monitoring and fault diagnosis in machine tools,” *Int. J. Mach. Tools Manuf.*, vol. 34, no. 4, pp. 527–551, 1994.

[2] J. Lee, R. Abujamra, A. K. S. Jardine, D. Lin, and D. Banjevic, “An integrated platform for diagnostics, prognostics and maintenance optimization,” *Proc. Intell. Maint. Syst.*, pp. 15–27, 2004.

[3] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[4] J. E. Berry, “How to track rolling element bearing health with vibration signature analysis,” *Sound Vib.*, vol. 25, no. 11, pp. 24–35, 1991.

[5] R. Moore, F. Pardue, A. Pride, and J. Wilson, “The Reliability-based maintenance strategy: a vision for improving industrial productivity,” *CSI Ind. Report, Knoxville, TN*, 1993.

[6] “What is a Bearing?” [Online]. Available: https://www.nskamericas.com/en/services/what-s-a-bearing.html. [Accessed: 01-Jun-2019].

[7] “Introduction to Bearings,” *2018, © NSK Ltd.* [Online]. Available: https://www.nsk.com/services/basicknowledge/introduction/index.html. [Accessed: 01-Jun-2019].

[8] “Structure and composition,” © *BearingsInChina*. [Online]. Available: http://www.bearingsinchina.com/structure-and-composition. [Accessed: 01-Jun-2019].

[9] K. N. HARRISTA, “Rolling bearing analysis: advanced concepts of bearing technology.” 5thed. NewYork: TaylorandFrancisGroup, 2007.

[10] “Bearing Failure: Causes and Cures,” *Schaeffler Group*. [Online]. Available: https://www.schaeffler.com/remotemedien/media/_shared_media/08_media_library/01_publications/barden/brochure_2/downloads_24/barden_bearing_failures_us_en.pdf. [Accessed: 01-Jun-2019].

[11] I. Howard, “A Review of Rolling Element Bearing Vibration’Detection, Diagnosis and Prognosis’,” DEFENCE SCIENCE AND TECHNOLOGY ORGANIZATION CANBERRA (AUSTRALIA), 1994.

[12] P. Y. Kim, “A review of rolling element bearing health monitoring. III- Preliminary test results on eddy current proximity transducer technique,” in *International Conference on Vibrations in Rotating Machinery, 3 rd, Heslington, England*, 1984, pp. 119–125.

[13] S. Teo, “Condition monitoring of slow speed rolling element bearings in a mechanically noisy environment,” *Appita J.*, vol. 42, no. 3, pp. 206–208, 1989.

[14] R. L. Smith, “Rolling Element Bearing Diagnostics with Lasers, Microphones and Accelerometers,” *Vib. Institute(USA),* pp. 43–52, 1992.

[15] J. R. Matthews, *Acoustic emission*, vol. 2. CRC Press, 1983.

[16] T. J. Holroyd and N. Randall, “Use of acoustic emission for machine condition monitoring,” *Br. J. Non-Destructive Test.*, vol. 35, no. 2, pp. 75–78, 1993.

[17] M. P. Norton and D. G. Karczub, *Fundamentals of noise and vibration analysis for engineers*. Cambridge university press, 2003.

[18] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011.

[19] S. S. Rao and F. F. Yap, *Mechanical vibrations*, vol. 4. Prentice Hall Upper Saddle River, 2011.

[20] A. V Dube, L. S. Dhamande, and P. G. Kulkarni, “Vibration based condition assessment of rolling element bearings with localized defects,” *Int. J. Sci. Technol. Res*, vol. 2, no. 4, pp. 149–155, 2013.

[21] L. Saidi, J. Ben Ali, and F. Fnaiech, “Bi-spectrum based-EMD applied to the non-stationary vibration signals for bearing faults diagnosis,” *ISA Trans.*, vol. 53, no. 5, pp. 1650–1660, 2014.

[22] R. X. Gao and R. Yan, “Non-stationary signal processing for bearing health monitoring,” *Int. J. Manuf. Res.*, vol. 1, no. 1, pp. 18–40, 2006.

[23] G. K. Singh, “Experimental investigations on induction machine condition monitoring and fault diagnosis using digital signal processing techniques,” *Electr. Power Syst. Res.*, vol. 65, no. 3, pp. 197–221, 2003.

[24] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*, vol. 31999. McGraw-Hill New York, 1986.

[25] J. Allen, “Short term spectral analysis, synthesis, and modification by discrete Fourier transform,” *IEEE Trans. Acoust.*, vol. 25, no. 3, pp. 235–238, 1977.

[26] L. Xiang and A. Hu, “Comparison of Methods for Different Time-frequency Analysis of Vibration Signal.,” *JSW*, vol. 7, no. 1, pp. 68–74, 2012.

[27] E. P. Wigner, “On the quantum correction for thermodynamic equilibrium,” in *Part I:*

Physical Chemistry. Part II: Solid State Physics, Springer, 1997, pp. 110–120.

[28] J. Antoni and R. B. Randall, "The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 308–331, 2006.

[29] N. Sawalhi and R. B. Randall, "Spectral kurtosis optimization for rolling element bearings.," in *ISSPA*, 2005, pp. 839–842.

[30] Y. Lei, J. Lin, Z. He, and Y. Zi, "Application of an improved kurtogram method for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1738–1749, 2011.

[31] N. Sawalhi, R. B. Randall, and H. Endo, "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2616–2633, 2007.

[32] D. Wang, W. T. Peter, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 35, no. 1–2, pp. 176–199, 2013.

[33] J. Antoni, F. Bonnardot, A. Raad, and M. El Badaoui, "Cyclostationary modelling of rotating machine vibration signals," *Mech. Syst. Signal Process.*, vol. 18, no. 6, pp. 1285–1314, 2004.

[34] C. Li, D. Cabrera, J. V. de Oliveira, R.-V. Sanchez, M. Cerrada, and G. Zurita, "Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram," *Mech. Syst. Signal Process.*, vol. 76, pp. 157–173, 2016.

[35] P. Borghesani, P. Pennacchi, R. Ricci, and S. Chatterton, "Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals," *Mech. Syst. Signal Process.*, vol. 40, no. 1, pp. 38–55, 2013.

[36] Y. Zhou, J. Chen, G. M. Dong, W. B. Xiao, and Z. Y. Wang, "Application of the horizontal slice of cyclic bispectrum in rolling element bearings diagnosis," *Mech. Syst. Signal Process.*, vol. 26, pp. 229–243, 2012.

[37] W. Cioch, O. Knapik, and J. Leśkow, "Finding a frequency signature for a cyclostationary signal with applications to wheel bearing diagnostics," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 55–64, 2013.

[38] A. Moshrefzadeh and A. Fasana, "The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis," *Mech. Syst. Signal Process.*, vol. 105, pp. 294–318, 2018.

[39] W. T. Peter, Y. H. and Peng, and R. Yam, "Wavelet analysis and envelope detection for

rolling element bearing fault diagnosis—their effectiveness and flexibilities," *J. Vib. Acoust.*, vol. 123, no. 3, pp. 303–310, 2001.

[40] V. Purushotham, S. Narayanan, and S. A. N. Prasad, "Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition," *Ndt E Int.*, vol. 38, no. 8, pp. 654–664, 2005.

[41] J. Lin and L. Qu, "Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis," *J. Sound Vib.*, vol. 234, no. 1, pp. 135–148, 2000.

[42] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine," *Mech. Syst. Signal Process.*, vol. 21, no. 7, pp. 2933–2945, 2007.

[43] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques," *Mech. Syst. Signal Process.*, vol. 25, no. 6, pp. 2083–2101, 2011.

[44] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 688–705, 2007.

[45] Y.-F. Li, M. Zuo, K. Feng, and Y.-J. Chen, "Detection of bearing faults using a novel adaptive morphological update lifting wavelet," *Chinese J. Mech. Eng.*, vol. 30, no. 6, pp. 1305–1313, 2017.

[46] S. Q. Zhang, N. He, J. T. Lv, X. H. Xu, and X. Y. Zang, "Research of the lifting wavelet arithmetic and applied in rotary mechanic fault diagnosis," in *Journal of physics: conference series*, 2006, vol. 48, no. 1, p. 696.

[47] A. Zifan, M. H. Moradi, and S. Gharibzadeh, "Microarray image enhancement by denoising using decimated and undecimated multiwavelet transforms," *Signal, Image Video Process.*, vol. 4, no. 2, pp. 177–185, 2010.

[48] J. Yuan, Z. He, Y. Zi, and Y. Wei, "Construction and selection of lifting-based multiwavelets for mechanical fault detection," *Mech. Syst. Signal Process.*, vol. 40, no. 2, pp. 571–588, 2013.

[49] X. Wang, Y. Zi, and Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 285–304, 2011.

[50] J. Yuan, Z. He, Y. Zi, Y. Lei, and Z. Li, "Adaptive multiwavelets via two-scale similarity transforms for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 23, no. 5, pp. 1490–1508, 2009.

[51] X. Wang, Y. Zi, and Z. He, “Multiwavelet construction via an adaptive symmetric lifting scheme and its applications for rotating machinery fault diagnosis,” *Meas. Sci. Technol.*, vol. 20, no. 4, p. 45103, 2009.

[52] J. Chen, M. J. Zuo, Y. Zi, and Z. He, “Construction of customized redundant multiwavelet via increasing multiplicity for fault detection of rotating machinery,” *Mech. Syst. Signal Process.*, vol. 42, no. 1–2, pp. 206–224, 2014.

[53] B. Walczak and D. L. Massart, “Noise suppression and signal compression using the wavelet packet transform,” *Chemom. Intell. Lab. Syst.*, vol. 36, no. 2, pp. 81–94, 1997.

[54] B. Liu, “Selection of wavelet packet basis for rotating machinery fault diagnosis,” *J. Sound Vib.*, vol. 284, no. 3–5, pp. 567–582, 2005.

[55] N. G. Nikolaou and I. A. Antoniadis, “Rolling element bearing fault diagnosis using wavelet packets,” *Ndt E Int.*, vol. 35, no. 3, pp. 197–205, 2002.

[56] C. Shen, D. Wang, F. Kong, and W. T. Peter, “Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier,” *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2013.

[57] L.-Y. Zhao, L. Wang, and R.-Q. Yan, “Rolling bearing fault diagnosis based on wavelet packet decomposition and multi-scale permutation entropy,” *Entropy*, vol. 17, no. 9, pp. 6447–6461, 2015.

[58] N. E. Huang *et al.*, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[59] Y. Yu and C. Junsheng, “A roller bearing fault diagnosis method based on EMD energy entropy and ANN,” *J. Sound Vib.*, vol. 294, no. 1–2, pp. 269–277, 2006.

[60] Y. Lei, Z. He, Y. Zi, and Q. Hu, “Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs,” *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2280–2294, 2007.

[61] Y. Yang, D. Yu, and J. Cheng, “A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM,” *Measurement*, vol. 40, no. 9–10, pp. 943–950, 2007.

[62] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, “Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals,” *Appl. Acoust.*, vol. 89, pp. 16–27, 2015.

[63] Y. Lei, Z. He, and Y. Zi, “A new approach to intelligent fault diagnosis of rotating machinery,” *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1593–1600, 2008.

[64] D. Yu, J. Cheng, and Y. Yang, “Application of EMD method and Hilbert spectrum to

the fault diagnosis of roller bearings," *Mech. Syst. Signal Process.*, vol. 19, no. 2, pp. 259–270, 2005.

[65] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 01, pp. 1–41, 2009.

[66] X. Zhang, Y. Liang, and J. Zhou, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.

[67] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mech. Syst. Signal Process.*, vol. 41, no. 1–2, pp. 127–140, 2013.

[68] A. Tabrizi, L. Garibaldi, A. Fasana, and S. Marchesiello, "Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine," *Meccanica*, vol. 50, no. 3, pp. 865–874, 2015.

[69] L. Meng, J. Xiang, Y. Wang, Y. Jiang, and H. Gao, "A hybrid fault diagnosis method using morphological filter–translation invariant wavelet and improved ensemble empirical mode decomposition," *Mech. Syst. Signal Process.*, vol. 50, pp. 101–115, 2015.

[70] L. Meng, J. Xiang, Y. Zhong, and W. Song, "Fault diagnosis of rolling bearing based on second generation wavelet denoising and morphological filter," *J. Mech. Sci. Technol.*, vol. 29, no. 8, pp. 3121–3129, 2015.

[71] J. Xiang, Y. Zhong, and H. Gao, "Rolling element bearing fault detection using PPCA and spectral kurtosis," *Measurement*, vol. 75, pp. 180–191, 2015.

[72] Z. Liu, Z. He, W. Guo, and Z. Tang, "A hybrid fault diagnosis method based on second generation wavelet de-noising and local mean decomposition for rotating machinery," *ISA Trans.*, vol. 61, pp. 211–220, 2016.

[73] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 647–656, 2015.

[74] L. Lu, J. Yan, and C. W. de Silva, "Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition," *J. Sound Vib.*, vol. 344, pp. 464–483, 2015.

[75] L. Saidi, J. Ben Ali, and F. Fnaiech, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA Trans.*, vol. 54, pp. 193–

206, 2015.

[76] Z. Su, B. Tang, Z. Liu, and Y. Qin, “Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine,” *Neurocomputing*, vol. 157, pp. 208–222, 2015.

[77] J. Ben Ali, L. Saidi, A. Mouelhi, B. Chebel-Morello, and F. Fnaiech, “Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations,” *Eng. Appl. Artif. Intell.*, vol. 42, pp. 67–81, 2015.

[78] X. Wang, Y. Zheng, Z. Zhao, and J. Wang, “Bearing fault diagnosis based on statistical locally linear embedding,” *Sensors*, vol. 15, no. 7, pp. 16225–16247, 2015.

[79] F. Li, J. Wang, M. K. Chyu, and B. Tang, “Weak fault diagnosis of rotating machinery based on feature reduction with Supervised Orthogonal Local Fisher Discriminant Analysis,” *Neurocomputing*, vol. 168, pp. 505–519, 2015.

[80] H. Zhou, J. Chen, G. Dong, H. Wang, and H. Yuan, “Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model,” *Mech. Syst. Signal Process.*, vol. 66, pp. 568–581, 2016.

[81] Z. Lv, B. Tang, Y. Zhou, and C. Zhou, “A novel method for mechanical fault diagnosis based on variational mode decomposition and multikernel support vector machine,” *Shock Vib.*, vol. 2016, 2016.

[82] “Bearing rating life.” [Online]. Available: https://www.skf.com/group/products/bearings-units-housings/principles/bearing-selection-process/bearing-size/size-selection-based-on-rating-life/bearing-rating-life/index.html. [Accessed: 01-Jun-2019].

[83] “Bearing Life.” [Online]. Available: http://www.dac-hvac.com/ask-rick-bearing-life-what-is-l10/. [Accessed: 01-Jun-2019].

[84] N. Li, Y. Lei, J. Lin, and S. X. Ding, “An improved exponential model for predicting remaining useful life of rolling element bearings,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, 2015.

[85] “Log-Normal distribution.” [Online]. Available: https://support.minitab.com/en-us/minitab/18/help-and-howto/probabilitydistributions-and-random-data/supporting-topics/distributions/lognormaldistribution/. [Accessed: 01-Jun-2019].

[86] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, “Rotating machinery prognostics: State of the art, challenges and opportunities,” *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.

[87] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Adaptive prognostics for rolling element bearing condition," *Mech. Syst. Signal Process.*, vol. 13, no. 1, pp. 103–113, 1999.

[88] Y. Qian, R. Yan, and R. X. Gao, "A multi-time scale approach to remaining useful life prediction in rolling bearing," *Mech. Syst. Signal Process.*, vol. 83, pp. 549–567, 2017.

[89] L. Saidi, J. Ben Ali, M. Benbouzid, and E. Bechhofer, "An integrated wind turbine failures prognostic approach implementing Kalman smoother with confidence bounds," *Appl. Acoust.*, vol. 138, pp. 199–208, 2018.

[90] C. H. Oppenheimer and K. A. Loparo, "Physically based diagnosis and prognosis of cracked rotor shafts," in *Component and Systems Diagnostics, Prognostics, and Health Management II*, 2002, vol. 4733, pp. 122–133.

[91] R. F. Orsagh, J. Sheldon, and C. J. Klenke, "Prognostics/diagnostics for gas turbine engine bearings," in *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*, 2003, vol. 7, pp. 3095–3103.

[92] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation–a review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.

[93] Z. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Appl. Stoch. Model. Bus. Ind.*, vol. 31, no. 1, pp. 16–32, 2015.

[94] Z. Omar, B. hmida Faycel, M. M. Hedi, and C. Abdelkader, "Stochastic Modeling of Wear in Bearing in Motor Pump in Two–Tank System," in *2018 15th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2018, pp. 611–618.

[95] Z.-S. Ye and N. Chen, "The inverse Gaussian process as a degradation model," *Technometrics*, vol. 56, no. 3, pp. 302–311, 2014.

[96] X.-S. Si, T. Li, Q. Zhang, and Z.-X. Zhang, "A prognostic model for degrading systems with randomly arriving shocks," in *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, 2016, pp. 1–4.

[97] Z.-X. Zhang, C.-H. Hu, X.-S. Si, and S.-H. Zhou, "A degradation-modeling based prognostic approach for systems with switching operating process," in *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, 2016, pp. 1–6.

[98] J.-X. Zhang, C.-H. Hu, X. He, X.-S. Si, Y. Liu, and D.-H. Zhou, "Lifetime prognostics for deteriorating systems with time-varying random jumps," *Reliab. Eng. Syst. Saf.*, vol. 167, pp. 338–350, 2017.

[99] B. Zhang, L. Xu, Y. Chen, and A. Li, "Remaining useful life based maintenance policy

for deteriorating systems subject to continuous degradation and shock," *Procedia CIRP*, vol. 72, pp. 1311–1315, 2018.

[100] D.-B. Du, J.-X. Zhang, Z.-J. Zhou, X.-S. Si, and C.-H. Hu, "Estimating remaining useful life for degrading systems with large fluctuations," *J. Control Sci. Eng.*, vol. 2018, 2018.

[101] Y. Hu *et al.*, "A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process," *Renew. energy*, vol. 127, pp. 452–460, 2018.

[102] J. Wen, H. Gao, and J. Zhang, "Bearing remaining useful life prediction based on a nonlinear wiener process model," *Shock Vib.*, vol. 2018, 2018.

[103] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Trans. Ind. Informatics*, vol. 12, no. 3, pp. 924–932, 2016.

[104] C. Ge, Y. Zhu, and Y. Di, "Hybrid Degradation Equipment Remaining Useful Life Prediction Oriented Parallel Simulation considering Model Soft Switch," *Comput. Intell. Neurosci.*, vol. 2019, 2019.

[105] P. Wang, R. Yan, and R. X. Gao, "Multi-Mode Particle Filter for Bearing Remaining Life Prediction," in *ASME 2018 13th International Manufacturing Science and Engineering Conference*, 2018, p. V003T02A031-V003T02A031.

[106] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Trans.*, vol. 45, no. 9, pp. 939–952, 2013.

[107] C. K. R. Lim and D. Mba, "Switching Kalman filter for failure prognostic," *Mech. Syst. Signal Process.*, vol. 52, pp. 426–435, 2015.

[108] Z.-Q. Wang, C.-H. Hu, and H.-D. Fan, "Real-time remaining useful life prediction for a nonlinear degrading system in service: Application to bearing data," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 211–222, 2017.

[109] J. Wu, C. Wu, S. Cao, S. W. Or, C. Deng, and X. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 529–539, 2018.

[110] H. Wang, Y. Zhao, and X. Ma, "Remaining Useful Life Prediction Using a Novel Two-Stage Wiener Process With Stage Correlation," *IEEE Access*, vol. 6, pp. 65227–65238, 2018.

[111] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models," *IEEE Trans. Reliab.*, vol. 61, no. 2, pp. 491–503, 2012.

[112] K. Medjaher, D. A. Tobon-Mejia, and N. Zerhouni, "Remaining useful life estimation of

critical components with application to bearings," *IEEE Trans. Reliab.*, vol. 61, no. 2, pp. 292–302, 2012.

[113] X. Zhang, R. Xu, C. Kwan, S. Y. Liang, Q. Xie, and L. Haynes, "An integrated approach to bearing fault diagnostics and prognostics," in *Proceedings of the 2005, American Control Conference, 2005.*, 2005, pp. 2750–2755.

[114] T. T. Le, C. Berenguer, and F. Chatelain, "Multi-branch hidden semi-markov modeling for rul prognosis," in *2015 Annual Reliability and Maintainability Symposium (RAMS)*, 2015, pp. 1–6.

[115] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman filtering for remaining-useful-life estimation of bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, 2014.

[116] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "An integrated Bayesian approach to prognositics of the remaining useful life and its application on bearing degradation problem," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, 2015, pp. 1090–1095.

[117] C. Ge, Y. Zhu, and Y. Di, "Equipment remaining useful life prediction oriented symbiotic simulation driven by real-time degradation data," *Int. J. Model. Simulation, Sci. Comput.*, vol. 9, no. 02, p. 1850009, 2018.

[118] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. S. Chow, "Anomaly detection and fault prognosis for bearings," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 2046–2054, 2016.

[119] L. C. K. Reuben and D. Mba, "Diagnostics and prognostics using switching Kalman filters," *Struct. Heal. Monit.*, vol. 13, no. 3, pp. 296–306, 2014.

[120] L. Cui, X. Wang, Y. Xu, H. Jiang, and J. Zhou, "A novel switching unscented Kalman filter method for remaining useful life prediction of rolling bearing," *Measurement*, vol. 135, pp. 678–684, 2019.

[121] Y. Qian and R. Yan, "Remaining useful life prediction of rolling bearings using an enhanced particle filter," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 10, pp. 2696–2707, 2015.

[122] J. Deutsch, M. He, and D. He, "Remaining useful life prediction of hybrid ceramic bearings using an integrated deep learning and particle filter approach," *Appl. Sci.*, vol. 7, no. 7, p. 649, 2017.

[123] K. Li, J. Wu, Q. Zhang, L. Su, and P. Chen, "New Particle Filter Based on GA for Equipment Remaining Useful Life Prediction," *Sensors*, vol. 17, no. 4, p. 696, 2017.

[124] T. Boukra, “Identifying new prognostic features for remaining useful life prediction using particle filtering and neuro-fuzzy system predictor,” in *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, 2015, pp. 1533–1538.

[125] X. Liu, P. Song, C. Yang, C. Hao, and W. Peng, “Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1549–1558, 2017.

[126] B. Fan, L. Hu, and N. Hu, “1551. Remaining useful life prediction of rolling bearings by the particle filter method based on degradation rate tracking.,” *J. Vibroengineering*, vol. 17, no. 2, 2015.

[127] T. H. Loutas, D. Roulias, and G. Georgoulas, “Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression,” *IEEE Trans. Reliab.*, vol. 62, no. 4, pp. 821–832, 2013.

[128] A. Soualhi, K. Medjaher, and N. Zerhouni, “Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, 2014.

[129] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, “Remaining useful life estimation based on nonlinear feature reduction and support vector regression,” *Eng. Appl. Artif. Intell.*, vol. 26, no. 7, pp. 1751–1760, 2013.

[130] E. Fumeo, L. Oneto, and D. Anguita, “Condition based maintenance in railway transportation systems based on big data streaming analysis,” *Procedia Comput. Sci.*, vol. 53, pp. 437–446, 2015.

[131] Z. Liu, M. J. Zuo, and Y. Qin, “Remaining useful life prediction of rolling element bearings based on health state assessment,” *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 230, no. 2, pp. 314–330, 2016.

[132] Z. Liu, Q. Li, X. Liu, and C. Mu, “A hybrid LSSVR/HMM-based prognostic approach,” *Sensors*, vol. 13, no. 5, pp. 5542–5560, 2013.

[133] A. Rai and S. H. Upadhyay, “Intelligent bearing performance degradation assessment and remaining useful life prediction based on self-organising map and support vector regression,” *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 232, no. 6, pp. 1118–1132, 2018.

[134] A. Rai and S. H. Upadhyay, “An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence,” *Appl. Soft Comput.*, vol. 71, pp. 36–50, 2018.

[135] L. Cao, Z. Qian, and Y. Pei, “Remaining Useful Life Prediction of Wind Turbine Generator Bearing Based on EMD with an Indicator,” in *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, 2018, pp. 375–379.

[136] V. N. Vladimir and V. Vapnik, “The nature of statistical learning theory.” Springer Heidelberg, 1995.

[137] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[138] R. Huang, L. Xi, X. Li, C. R. Liu, H. Qiu, and J. Lee, “Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods,” *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 193–207, 2007.

[139] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, “A recurrent neural network based health indicator for remaining useful life prediction of bearings,” *Neurocomputing*, vol. 240, pp. 98–109, 2017.

[140] Z. Tian, “An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring,” *J. Intell. Manuf.*, vol. 23, no. 2, pp. 227–237, 2012.

[141] A. K. Mahamad, S. Saon, and T. Hiyama, “Predicting remaining useful life of rotating machinery based artificial neural network,” *Comput. Math. with Appl.*, vol. 60, no. 4, pp. 1078–1087, 2010.

[142] L. Ren, J. Cui, Y. Sun, and X. Cheng, “Multi-bearing remaining useful life collaborative prediction: A deep learning approach,” *J. Manuf. Syst.*, vol. 43, pp. 248–256, 2017.

[143] L. Ren, Y. Sun, H. Wang, and L. Zhang, “Prediction of bearing remaining useful life with deep convolution neural network,” *IEEE Access*, vol. 6, pp. 13041–13049, 2018.

[144] A. Rai and S. H. Upadhyay, “The use of MD-CUMSUM and NARX neural network for anticipating the remaining useful life of bearings,” *Measurement*, vol. 111, pp. 397–410, 2017.

[145] B. Zhang, S. Zhang, and W. Li, “Bearing performance degradation assessment using long short-term memory recurrent network,” *Comput. Ind.*, vol. 106, pp. 14–29, 2019.

[146] A. Z. Hinchi and M. Tkiouat, “Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network,” *Procedia Comput. Sci.*, vol. 127, pp. 123–132, 2018.

[147] J. Zhu, N. Chen, and W. Peng, “Estimation of bearing remaining useful life based on multiscale convolutional neural network,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, 2018.

[148] W. Ahmad, S. Ali Khan, and J.-M. Kim, “Estimating the remaining useful life of

bearings using a neuro-local linear estimator-based method," *J. Acoust. Soc. Am.*, vol. 141, no. 5, pp. EL452–EL457, 2017.

[149] Z. Tian, "An artificial neural network approach for remaining useful life prediction of equipments subject to condition monitoring," in *2009 8th International Conference on Reliability, Maintainability and Safety*, 2009, pp. 143–148.

[150] L. Xiao, X. Chen, X. Zhang, and M. Liu, "A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition," *J. Intell. Manuf.*, vol. 28, no. 8, pp. 1893–1914, 2017.

[151] J. Cui, L. Ren, X. Wang, and L. Zhang, "Pairwise comparison learning based bearing health quantitative modeling and its application in service life prediction," *Futur. Gener. Comput. Syst.*, 2019.

[152] W. Mao, J. He, J. Tang, and Y. Li, "Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network," *Adv. Mech. Eng.*, vol. 10, no. 12, p. 1687814018817184, 2018.

[153] S. A. Khan, A. E. Prosvirin, and J.-M. Kim, "Towards bearing health prognosis using generative adversarial networks: Modeling bearing degradation," in *2018 International Conference on Advancements in Computational Sciences (ICACS)*, 2018, pp. 1–6.

[154] J. S. L. Senanayaka, H. Van Khang, and K. G. Robbersmyr, "Autoencoders and Recurrent Neural Networks Based Algorithm for Prognosis of Bearing Life," in *2018 21st International Conference on Electrical Machines and Systems (ICEMS)*, 2018, pp. 537–542.

[155] Z. Liu and Y. Guo, "A neural network approach for prediction of bearing performance degradation tendency," in *2017 9th International Conference on Modelling, Identification and Control (ICMIC)*, 2017, pp. 204–208.

[156] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. energy Convers.*, vol. 20, no. 4, pp. 719–729, 2005.

[157] J. Ben Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mech. Syst. Signal Process.*, vol. 56, pp. 150–172, 2015.

[158] B. Satish and N. D. R. Sarma, "A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors," in *IEEE Power Engineering Society General Meeting, 2005*, 2005, pp. 2291–2294.

[159] Z. Cheng and B. Cai, "Predicting the remaining useful life of rolling element bearings

using locally linear fusion regression," *J. Intell. Fuzzy Syst.*, no. Preprint, pp. 1–12, 2018.

[160] X. Li, "Remaining Useful Life Prediction of Bearings Using Fuzzy Multimodal Extreme Learning Regression," in *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, 2017, pp. 499–503.

[161] D. Zurita, J. A. Carino, M. Delgado, and J. A. Ortega, "Distributed neuro-fuzzy feature forecasting approach for condition monitoring," in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, 2014, pp. 1–8.

[162] A. C. Rencher, *Methods of multivariate analysis*, vol. 492. John Wiley & Sons, 2003.

[163] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data mining in time series databases*, World Scientific, 2004, pp. 1–21.

[164] L. Ma, J. S. Kang, and C. Y. Zhao, "Research on condition monitoring of bearing health using vibration data," in *Applied Mechanics and Materials*, 2012, vol. 226, pp. 340–344.

[165] R. Babuška and H. B. Verbruggen, "Constructing fuzzy models by product space clustering," in *Fuzzy model identification*, Springer, 1997, pp. 53–90.

[166] M. Sugeno, "An introductory survey of fuzzy control," *Inf. Sci. (Ny).*, vol. 36, no. 1–2, pp. 59–83, 1985.

[167] Fei Huang, Alexandre Sava, Kondo H Adjallah, and Zhouhang Wang, "Remaining Useful Life Estimation for Bearings Based on Segmented Projection Error and Fuzzy Inference System," in *The 6th International Conference on Mechanical Engineering & Mechanics*, 2017.

[168] F. Huang, A. Sava, K. H. Adjallah, and Z. Wang, "Bearings Degradation Monitoring Indicator Based on Segmented Hotelling T Square and Piecewise Linear Representation," in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2018, pp. 1389–1394.

[169] F. Huang, A. Sava, K. H. Adjallah, and Z. Wang, "Data Piecewise Linear Approximation for Bearings Degradation Monitoring," in *The 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 2019, pp. 60–64.

[170] E. Keogh, S. Chu, and M. Pazzani, "Ensemble-index: A new approach to indexing large databases," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 117–125.

[171] D. Rafiei and A. Mendelzon, "Similarity-based queries for time series data," in *ACM SIGMOD Record*, 1997, vol. 26, no. 2, pp. 13–25.

[172] I. Popivanov and R. J. Miller, "Similarity search over time-series data using wavelets,"

in *Proceedings 18th international conference on data engineering*, 2002, pp. 212–221.

[173] C. Shahabi, X. Tian, and W. Zhao, "Tsa-tree: A wavelet-based approach to improve the efficiency of multi-level surprise and trend queries on time-series data," in *Proceedings. 12th International Conference on Scientific and Statistica Database Management*, 2000, pp. 55–68.

[174] K. Kawagoe and T. Ueda, "A similarity search method of time series data with combination of fourier and wavelet transforms," in *Proceedings Ninth International Symposium on Temporal Representation and Reasoning*, 2002, pp. 86–92.

[175] K. J. Åström, "On the choice of sampling rates in parametric identification of time series," *Inf. Sci. (Ny).*, vol. 1, no. 3, pp. 273–278, 1969.

[176] B.-K. Yi and C. Faloutsos, "Fast time sequence indexing for arbitrary Lp norms," in *VLDB*, 2000, vol. 385, no. 394, p. 99.

[177] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowl. Inf. Syst.*, vol. 3, no. 3, pp. 263–286, 2001.

[178] S. Lee, D. Kwon, and S. Lee, "Dimensionality reduction for indexing time series based on the minimum distance," *J. Inf. Sci. Eng.*, vol. 19, no. 4, pp. 697–711, 2003.

[179] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Locally adaptive dimensionality reduction for indexing large time series databases," *ACM Sigmod Rec.*, vol. 30, no. 2, pp. 151–162, 2001.

[180] H. Shatkay and S. B. Zdonik, "Approximate queries and representations for large data sequences," in *Proceedings of the Twelfth International Conference on Data Engineering*, 1996, pp. 536–545.

[181] E. J. Keogh and M. J. Pazzani, "An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback.," in *Kdd*, 1998, vol. 98, pp. 239–243.

[182] S. Park, D. Lee, and W. W. Chu, "Fast retrieval of similar subsequences in long sequence databases," in *Proceedings 1999 Workshop on Knowledge and Data Engineering Exchange (KDEX'99)(Cat. No. PR00453)*, 1999, pp. 60–67.

[183] J. Hunter and N. McIntosh, "Knowledge-based event detection in complex time series data," in *Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making*, 1999, pp. 271–280.

[184] E. J. Keogh and P. Smyth, "A probabilistic approach to fast pattern matching in time series databases.," in *Kdd*, 1997, vol. 1997, pp. 24–30.

[185] M. Ishijima, S.-B. Shin, G. H. Hostetter, and J. Sklansky, “Scan-along polygonal approximation for data compression of electrocardiograms,” *IEEE Trans. Biomed. Eng.*, no. 11, pp. 723–729, 1983.

[186] H. Vullings, M. H. G. Verhaegen, and H. B. Verbruggen, “ECG segmentation using time-warping,” in *International Symposium on Intelligent Data Analysis*, 1997, pp. 275–285.

[187] V. Guralnik and J. Srivastava, “Event detection from time series data,” in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 1999, pp. 33–42.

[188] P. Nomikos and J. F. MacGregor, “Multivariate SPC charts for monitoring batch processes,” *Technometrics*, vol. 37, no. 1, pp. 41–59, 1995.

[189] M. A. Kramer, “Nonlinear principal component analysis using autoassociative neural networks,” *AIChE J.*, vol. 37, no. 2, pp. 233–243, 1991.

[190] J. Chen and K.-C. Liu, “On-line batch process monitoring using dynamic PCA and dynamic PLS models,” *Chem. Eng. Sci.*, vol. 57, no. 1, pp. 63–75, 2002.

[191] S. Rännar, J. F. MacGregor, and S. Wold, “Adaptive batch monitoring using hierarchical PCA,” *Chemom. Intell. Lab. Syst.*, vol. 41, no. 1, pp. 73–81, 1998.

[192] S. Valle, W. Li, and S. J. Qin, “Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods,” *Ind. Eng. Chem. Res.*, vol. 38, no. 11, pp. 4389–4401, 1999.

[193] H. Hotelling, “Analysis of a complex of statistical variables into principal components.,” *J. Educ. Psychol.*, vol. 24, no. 6, p. 417, 1933.

[194] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, “The mahalanobis distance,” *Chemom. Intell. Lab. Syst.*, vol. 50, no. 1, pp. 1–18, 2000.

[195] E. Keogh, S. Chu, D. Hart, and M. Pazzani, “An online algorithm for segmenting time series,” in *ICDM 2001, Proceedings IEEE International Conference on Data Mining*, 2001, pp. 289–296.

[196] D. J. Best and D. E. Roberts, “Algorithm AS 89: the upper tail probabilities of Spearman’s rho,” *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 24, no. 3, pp. 377–379, 1975.

[197] N. D. Singpurwalla and J. M. Booker, “Membership functions and probability measures of fuzzy sets,” *J. Am. Stat. Assoc.*, vol. 99, no. 467, pp. 867–877, 2004.

[198] L. A. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[199] S. Guillaume, “Designing fuzzy inference systems from data: An interpretability-oriented review,” *IEEE Trans. fuzzy Syst.*, vol. 9, no. 3, pp. 426–443, 2001.

[200] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” *Int. J. Man. Mach. Stud.*, vol. 7, no. 1, pp. 1–13, 1975.

[201] R. Murray-Smith and T. Johansen, *Multiple model approaches to nonlinear modelling and control*. CRC press, 1997.

[202] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Trans. Syst. Man. Cybern.*, no. 1, pp. 116–132, 1985.

[203] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[204] K. M. Bataineh, M. Naji, and M. Saqer, “A Comparison Study between Various Fuzzy Clustering Algorithms.,” *Jordan J. Mech. Ind. Eng.*, vol. 5, no. 4, 2011.

[205] J. C. Dunn, “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,” *Taylor Fr.*, 1973.

[206] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[207] R. R. Yager and D. P. Filev, “Generation of fuzzy rules by mountain clustering,” *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 209–219, 1994.

[208] S. L. Chiu, “Fuzzy model identification based on cluster estimation,” *J. Intell. fuzzy Syst.*, vol. 2, no. 3, pp. 267–278, 1994.

[209] J. H. Wolfe, “Pattern clustering by multivariate mixture analysis,” *Multivariate Behav. Res.*, vol. 5, no. 3, pp. 329–350, 1970.

[210] J.-S. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” *IEEE Trans. Syst. Man. Cybern.*, vol. 23, no. 3, pp. 665–685, 1993.

[211] R. MacAusland, “The Moore-Penrose Inverse and Least Squares,” *Math 420 Adv. Top. Linear Algebr.*, pp. 1–10, 2014.

[212] M. S. Wisz *et al.*, “Effects of sample size on the performance of species distribution models,” *Divers. Distrib.*, vol. 14, no. 5, pp. 763–773, 2008.

[213] S. Kaczmarz, “Approximate solution of systems of linear equations,” *Int. J. Control*, vol. 57, no. 6, pp. 1269–1271, 1993.

[214] F. Kovács, C. Legány, and A. Babos, “Cluster validity measurement techniques,” in *6th International symposium of hungarian researchers on computational intelligence*, 2005, p. 35.

[215] J. C. Dunn, “Well-separated clusters and optimal fuzzy partitions,” *J. Cybern.*, vol. 4, no. 1, pp. 95–104, 1974.

[216] R. W. Gunderson, “Application of fuzzy ISODATA algorithms to star tracker pointing

systems," *IFAC Proc. Vol.*, vol. 11, no. 1, pp. 1319–1323, 1978.

[217] M. P. Windham, "Cluster validity for fuzzy clustering algorithms," *Fuzzy Sets Syst.*, vol. 5, no. 2, pp. 177–185, 1981.

[218] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," *J. Math. Biol.*, vol. 1, no. 1, pp. 57–71, 1974.

[219] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 8, pp. 841–847, 1991.

[220] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 7, pp. 773–780, 1989.

[221] A. M. Bensaid *et al.*, "Validity-guided (re)clustering with applications to image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 112–123, 2002.

[222] S. Nefti, M. Oussalah, and U. Kaymak, "A New Fuzzy Set Merging Technique Using Inclusion-Based Fuzzy Clustering," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 145–161, 2008.

[223] A. Dourado, L. Aires, and J. V. Ramos, "eFSLab: Developing evolving fuzzy systems from data in a friendly environment," in *European Control Conference (ECC)*, 2009, pp. 922–927.

[224] C. R. Rao, "Advanced statistical methods in biometric research," *New York*, 1952.

[225] Å. Björck, T. Elfving, and S. Kaczmarz, "Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations," *BIT Numer. Math.*, vol. 57, no. 6, pp. 1269–1271, 1979.

[226] P. Nectoux *et al.*, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests.," in *IEEE International Conference on Prognostics and Health Management, PHM'12.*, 2012, pp. 1–8.

[227] H. Hellendoorn and D. Driankov, *Fuzzy model identification: selected approaches*. Springer Science & Business Media, 2012.

[228] D. E. Goldberg, "Genetic algorithms in search," *Optim. Mach.*, 1989.

[229] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science (80-.).*, vol. 220, no. 4598, pp. 671–680, 1983.

[230] Y. N. Pan, J. Chen, and X. L. Li, "Spectral entropy: a complementary index for rolling element bearing performance degradation assessment," *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 223, no. 5, pp. 1223–1231, 2009.

[231] R. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 824–839, 2007.

[232] W. Caesarendra, B. Kosasih, A. K. Tieu, and C. A. S. Moodie, “Application of the largest Lyapunov exponent algorithm for feature extraction in low speed slew bearing condition monitoring,” *Mech. Syst. Signal Process.*, vol. 50, pp. 116–138, 2015.

[233] D. Logan and J. Mathew, “Using the correlation dimension for vibration fault diagnosis of rolling element bearings—I. Basic concepts,” *Mech. Syst. Signal Process.*, vol. 10, no. 3, pp. 241–250, 1996.

[234] W. Caesarendra and T. Tjahjowidodo, “A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing,” *Machines*, vol. 5, no. 4, p. 21, 2017.

Résumé en français

Les travaux de recherche élaborés dans le cadre de cette thèse de doctorat se focalise sur l'estimation de la durée de vie résiduelle des roulements dans le cas où un volume de données historiques est disponible. Ces travaux sont motivés par le rôle important que les roulements jouent pour la performance des équipements de production rotatifs. La durée de vie résiduelle est un indicateur important pour la mise en place de stratégies de maintenance conditionnelle, basées sur l'état des équipements, dont l'objectif est de garantir un niveau de fiabilité adapté durant la vie utile de l'équipement.

La mise en place d'un programme de maintenance conditionnelle comporte trois étapes : collecte de données, analyse de données et prise de décision.

La collecte de données vise à accumuler et à stocker les données nécessaires pour extraire des informations utiles analyser et comprendre le processus de dégradation et pour caractériser de l'état de santé de l'équipement. Le processus de prise de décision en maintenance vise à établir une stratégie de maintenance efficace pour préserver l'aptitude de l'équipement à satisfaire ses missions.

Dans cette thèse nous focalisons notre attention sur le processus d'extraction des signatures et d'analyse afin d'estimer la durée de vie résiduelle des roulements à base des signaux vibratoires. Les signatures habituelles extraites des signaux de vibrations sont sensibles surtout dans le dernier stage du processus de dégradation. Nous avons proposé des nouvelles signatures, qui ont la particularité d'être monotones et qui incluent l'information historique sur le processus de dégradation.

Ensuite, nous avons élaboré une méthode basée sur un système d'inférence floue T-S (T-S FIS) pour l'estimation du RUL. Dans nos travaux nous avons utilisé les signaux collectées pour un petit nombre de roulements identiques durant un cycle de vie complet. Les données d'apprentissage utilisées pour l'identification des paramètres du modèle ont étés collectées suite à des observations périodiques. Nous avons développé une technique d'identification des paramètres du modèle T-S FIS basée sur l'analyse d'un mixage des distributions. Par la suite, nous avons utilisé la méthode du maximum de vraisemblance pour calcules les paramètres des clusters temporels et pour estimer la probabilité associée à chaque état de dégradation. Une méthode d'estimation basée sur les moindres carrées est utilisée pour identifier les paramètres de sortie du modèle.

Le processus de dégradation des roulements est complexe leur durée de vie présente une dispersion importante même pour des roulements identiques, soumis à des charges identiques. L'utilisation d'un ensemble de données d'apprentissage de petite taille ne permettant pas d'estimer les paramètres du modèle avec une grande précision, les paramètres du modèle doivent être améliorés lorsque des données supplémentaires deviennent disponibles. Par conséquent, nous avons élaboré une méthode pour la mise à jour des paramètres du modèle lorsque des nouvelles connaissances sont rendues disponibles suite à un processus de collecte de données adaptatif.

Cette méthode utilise le paramétrage initial du modèle ainsi que les nouveaux données d'apprentissage. Nous avons élaboré un algorithme itératif basé sur la méthode du maximum de vraisemblance pour la mise à jour des paramètres du modèle T-S FIS.

Des données fournies par un banc de test dans la littérature ont étés utilisées pour évaluer les méthodes proposées. Les résultats numériques ont montré la tendance monotone croissante des indicateurs proposés avec l'avancement de la dégradation. De même, les méthode d'identification et de mise à jour des paramètres ont étés comparées à des méthodes existantes dans la littérature, avec des résultats prometteurs.

Thesis summary in English

The research work elaborated in this thesis focuses on remaining useful life (RUL) estimation for bearings where poor historical training data is available. This research work is motivated by the importance of bearings RUL estimation for decision making in condition based maintenance framework of rotating machines in manufacturing systems. This maintenance strategy aims to define setup maintenance actions based on the health of equipment to ensure adequate reliability along with the useful life of facilities.

There are three key steps in a CBM program: data acquisition, data processing and maintenance decision-making.

The data acquisition step aims to collect and store the data relevant to the system health. Information is extracted form data during the data processing step in order to understand and analyze the degradation process and the condition of the equipment. The maintenance decision-making step aims to recommend efficient maintenance policies achieve the performance objectives assigned to the equipment.

In this work we focus on data acquisition and data processing for estimating the RUL based on vibration signals.

Common vibration signal based features for bearings degradation monitoring are sensible on the last stage of the degradation process. We propose new bearing degradation monitoring indicators that are monotonic and incorporate historical degradation information.

Furthermore, we proposed a Takagi - Sugeno fuzzy inference system (T-S FIS) based method for RUL estimation. In this work, we used the vibration signals collected from a small number of bearings over an entire period of run-to-failure. The training set is built using features extracted from periodic observations of the vibration signals. We elaborated a technique for T-S FIS parameters identification using a small size training dataset based on a mixture distribution analysis. The number of rules and the input parameters of each rule are identified using the subtractive clustering method. We use the maximum likelihood method of mixture distribution analysis to calculate the parameters of clusters on the time axis and the probability corresponding to each degradation stage. Based on this result, we identified the output parameters of each rule using a weighted least square estimation.

Bearings degradation process is complex and usually bearing lifetime is characterized by high dispersion, even for identical bearings under identical operating condition. A small size training set does not allow model parameters identification with high precision. To offset the drawback

of small size sample for model identification, in this work we propose a mixture distribution analysis based method for tuning the parameters of the T-S FIS model when additional knowledge becomes available. First, the input parameters of T-S FIS model are updated by using an iterative maximum likelihood estimation of mixture distribution analysis based method which is in this chapter. Then, the output parameters of T-S FIS model are updated by using another maximum likelihood estimation of mixture distribution analysis based method which is proposed in this chapter by using symmetrical Kaczmarz method. Finally, the tuned T-S FIS model is formed by using the updated parameters.

A benchmark is employed to evaluate the work proposed. The results show monotonic trend during the bearing degradation evolution for the indicators elaborated in this work. These indicators are sensitive to the bearing degradation process during the bearings entire lifecycle. The numerical results obtained for the model identification and the tuning methods proposed in this work provide good accuracy in case of small size additional data sets compared with existed methods. The method proposed in this work is promising for bearing RUL estimation when poor historical data is available.