WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY

DOCTORAL THESIS

Modelling and analysis of long-term historical data of time-varying complex systems in the presence of impulsive noise for condition monitoring

Author:
Hamid SHIRI

Supervisors:
Prof. dr hab. Radosław ZIMROZ
Dr hab. Agnieszka Wyłomańska, prof. uczelni

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the

Environmental engineering and Mining and Geology
Faculty of Geoengineering, Mining and Geology

July 8, 2024

"Prediction is very difficult, especially if it's about the future."

Nils Bohr

WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY

Abstract

Environmental engineering and Mining and Geology
Faculty of Geoengineering, Mining and Geology

Doctor of Philosophy

Modelling and analysis of long-term historical data of time-varying complex systems in the presence of impulsive noise for condition monitoring

by Hamid SHIRI

The field of long-term data analytics has gained considerable interest in recent years due to its crucial role in extracting useful insights, knowledge, and wisdom from long term datasets. In industrial contexts, the development of sensor technology and improvements in information and communication technologies have made it possible to collect large amounts of data, ranging from minutes to years. These carefully selected datasets are extremely helpful resources for informing and improving decision-making processes.
Using long-term data for machine health assessments and prognostics of machine is an important application in the field of condition-based maintenance (CBM). Early identification of system degradation and precise estimation of remaining useful life (RUL) are crucial for preserving the reliability and safety of industrial systems, while also reducing the risks associated with unexpected failures and maintenance expenses.

Complexities in long-term health index data, including non-stationary behavior, non-linearity, and non-Gaussian noise, pose severe challenges for machine health assessments and prognostic applications. The presence of non-stationary behavior in time series can lead to biased parameter estimates, erroneous correlations, and inaccurate predictions due to dynamic changes. This means that such data requires an advanced analysis approach and a complex model. On the other hand, non-Gaussian noise, especially heavy-tailed noise, breaks the Gaussian assumptions in time series models. This can cause estimates to be biased and make them more sensitive to outliers, so robust modeling techniques are needed for accurate analysis and prediction.

The thesis proposes a three-stage model to generate long-term health index degradation data, addressing non-stationary behavior and it's non-Gaussian character. It develops a robust framework using techniques capable of handling non-Gaussian noise to analyze and characterize historical degradation data. Additionally, it introduces robust offline and online segmentation methods based on deterministic trends in the health index, followed by the development of a maximum correntropy extended Kalman filter (MCEKF) for probabilistic estimation of RUL, considering the presence of non-Gaussian noise. Analytical studies on synthetic and real datasets with varying levels of non-Gaussian noise demonstrate the effectiveness of these approaches.

Keywords : machine health assessments, prognostics, segmentation, health index, remaining useful life, extended Kalman filter, robust methods, non-Gaussian noise, non-stationary.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors. Professor Radoslaw Zimroz has been more than just a supervisor; he has been like a father to me, offering support and guidance through all aspects of my life over these three years. Professor Agnieszka Wyłomańska's invaluable guidance, support, and encouragement have been essential to the completion of this thesis. Their insightful feedback and unwavering commitment have significantly enriched my research experience.

I am deeply grateful to my colleagues at the Digital Mining Center (DMC) for their collaboration, assistance, and camaraderie throughout this journey. In particular, I would like to express my heartfelt thanks to Pawel Zimroz, whose expertise, constructive criticism, and motivational words have been instrumental in shaping the direction and quality of my work.

In addition, I owe a special thanks to my beloved wife, whose love, patience, and unwavering support have been my anchor during the challenging moments of this endeavor. Her understanding and encouragement have been indispensable and I could not have achieved this milestone without her by my side.

To all those who have contributed to my personal and academic growth, directly or indirectly, I extend my sincere thanks. Your contributions have made this achievement possible and I am deeply grateful for your support.

Also, I gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie programme through the ETN MOIRA project (GA 955681).

Thank you.

Contents

Abstract **iii**

Acknowledgements **v**

List of Figures **xi**

List of Tables **xvii**

1 Introduction **1**
1.1 Evolvement of maintenance strategy 1
1.2 Condition based maintenance 2
1.3 Machinery's long term condition monitoring data 4

2 Problem formulation and state of the art **5**
2.1 Problem formulation 5
2.2 Purpose and objectives 6
2.3 Research questions 7
2.4 Author's Contribution to the Field 7
2.5 Authorship of mentioned papers 8
2.6 Scope of research 8
2.7 Limitations of this research 9
2.8 State of the art 9
2.8.1 Modeling of long-term health index data 9
2.8.2 Segmentation of long-term health index data 16

3 Experimental data **19**
3.1 Experimental data 19
3.1.1 FEMTO dataset 19
3.1.2 IMS dataset 21
3.1.3 Wind turbine dataset 22
3.2 Synthetic degradation model 23

4 Long-term health index data modeling and identification **29**
4.1 Introduction of long-term health index data modeling 29
4.2 Methodology: a framework for long-term health index data modelling 29
4.2.1 Algorithms for Gaussian distributed data 32
4.2.2 Algorithms for non-Gaussian heavy-tailed distributed data (robust approach) 33
4.2.3 Summary 34
4.3 Validation of the proposed approach by simulated data analysis 35
4.4 Real data analysis 41
4.4.1 Results of applying the proposed framework to FEMTO dataset 41
4.4.2 Results of applying the proposed framework to wind turbine dataset 42

4.4.3 Discussion . . . 44
4.5 Chapter highlights . . . 47

5 Segmentation of health index data **49**
5.1 Segmentation of long-term historical health index data . . . 49
5.1.1 Methodology of segmenting long-term historical health index data . . . 49
5.1.2 Piece-wise regression models . . . 49
Ordinary least squares . . . 50
Dynamic programming segmentation . . . 50
Iterative reweighted least squares . . . 50
Least absolute error . . . 51
5.1.3 Student's t distribution estimation . . . 51
5.1.4 Hidden Markov models and Expectation-Maximization algorithm . . . 52
Hidden Markov model regression . . . 52
Hidden logistic process . . . 53
5.1.5 Analysis of simulated data . . . 53
Analysis for Gaussian distributed model . . . 54
Analysis for non-Gaussian distributed model . . . 56
5.1.6 Real data analysis . . . 58
Results for IMS dataset . . . 58
Results for FEMTO dataset . . . 59
Results for wind turbine dataset . . . 60
Discussion . . . 61
5.2 Probabilistic segmentation of long-term health index data . . . 62
5.2.1 Methodology of probabilistic segmentation of long-term health index data . . . 62
Kalman filter . . . 62
Maximum correntropy Kalman filter . . . 64
Switching maximum correntropy Kalman filter . . . 65
5.2.2 Simulation . . . 69
Result for Gaussian distributed model . . . 69
Result for non-Gaussian distributed model . . . 70
5.2.3 Real data analysis . . . 72
Result for FEMTO dataset . . . 73
Results for Wind turbine dataset . . . 73
5.3 Chapter highlights . . . 74

6 Predicting of remaining useful life based on long-term health index **77**
6.1 Probabilistic methodology for predicting remaining useful life . . . 77
6.1.1 Remaining useful life . . . 78
6.1.2 State space degradation model . . . 79
6.1.3 Extended Kalman filter . . . 80
6.1.4 Discussion of the hyperparameters setting . . . 80
6.1.5 End of the life of machine's equipment . . . 81
6.2 Simulation and results . . . 82
6.2.1 Results for simulated signal . . . 82
Degradation process in the presence of the Gaussian noise . . . 82
Degradation process in the presence of the non-Gaussian noise . . . 84

6.3 Result for real dataset 87
6.3.1 Results for FEMTO dataset 88
6.3.2 Results for IMS dataset 90
6.3.3 Results for wind turbine dataset 92
6.3.4 Discussion 94

7 Conclusion, contributions and future research **97**
7.1 Conclusion 97
7.2 Practical perspectives on the proposed approaches 99
7.3 Contributions 100
7.4 Future research 100

A Considered distributions **103**

List of Figures

1.1 Maintenance strategy. 2
1.2 Condition based maintenance workflow. 3
1.3 Recorded vertical vibration of degraded bearing over time. 4
1.4 Recorded horizontal vibration of degraded bearing over time. 4

2.1 Helicopter noise source [3]. 5
2.2 Wind turbine health index. 6
2.3 Long-term degradation modeling, prediction approaches. 10
2.4 Statistical based-model approaches according the assumption of the noise. 12

3.1 FEMTO test rigs [97]. 20
3.2 Different faults of rolling element bearings [97]. 20
3.3 Raw vibration data for case study bearing $1-1$, each set of color is related a set of raw vibration data. 20
3.4 IMS test rig [114]. 21
3.5 Different faults of rolling element bearings [114]. 21
3.6 Raw vibration data for IMS dataset: subset 3, bearing 3, each set of color is related a set of raw vibration data. 22
3.7 Wind turbine test rigs [122]. 23
3.8 Wind turbine health index [122]. 23
3.9 Types of feature variations and degradation models: (a) good condition (constant trend), (b) good to gradual wear (linear trend) (c) switch from good to bad condition, (d) good to accelerated wear (linear trend), (e) good to accelerated wear (exponential trend), (f) three regimes model (good, linear progress and exponential progress of degradation). 24
3.10 The preliminary model used for our analysis. The red solid line represents the theoretical trend line, two vertical dashed lines point out the change of regime (Healthy Stage/ Degradation/ Critical stage). Note constant variance of random component. 24
3.11 Long-term data variation simulated from the adopted model. 25
3.12 The deterministic component $T(t)$ and the scale function $SC(t)$ for the following values of the parameters of the model: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$ and $c_1 = 10$. The left panel is shown the determinstic part of simulated HI and the right panel is demonstrated the scale of the random part. 27
3.13 The exemplary data from the model with deterministic component $T(t)$ and scale function $SC(t)$ presented in Fig. 3.12. The random components of the model correspond to the AR(1) model with $\phi = 0.5$. Four cases of distributions are considered: Gaussian (α-stable with $\alpha = 2$), and non-Gaussian α-stable with $\alpha = 1.95, 1.9$ and $\alpha = 1.85$. . . . 27

4.1 The framework for modelling of long-term data. Each block of this diagram is described in details in Section 4.2. 30
4.2 The outcomes of the proposed procedures for simulated data are organized across four rows, representing trend, scale, AR model, and distribution tail identification, respectively. These are presented in three columns: the left column indicates the classical approach tailored for Gaussian distributed signals, the middle column represents the robust approach suited for non-Gaussian heavy-tailed distributed signals, and the right column illustrates the evaluation of results through RMSE for trend and scale, alongside boxplots for random components. Specifically, the random component $R3(t)$ is distributed as Gaussian (i.e., α-stable with $\alpha = 2$). 37
4.3 The outcomes of the proposed procedures for simulated data are delineated across four rows, representing trend, scale, AR model, and distribution tail identification, respectively. The comparison is presented in three columns: the left column signifies the classical approach tailored for Gaussian distributed signals, the middle column represents the robust approach suited for non-Gaussian heavy-tailed distributed signals, and the right column illustrates the evaluation of results via RMSE for trend and scale, alongside boxplots for random components. Specifically, the random component $R3(t)$ is distributed according to the α-stable distribution with $\alpha = 1.95$. 38
4.4 The outcomes of the suggested methods for simulated data are presented in rows 1-4, representing trend, scale, AR model, and distribution tail identification, respectively. The left column is dedicated to the classical technique, which is suitable for signals that follow a Gaussian distribution. The middle column is dedicated to the robust approach, which is designed for signals that have a non-Gaussian heavy-tailed distribution. The right column evaluates the results using the RMSE for both the trend and scale, as well as boxplots for the random components. The random component $\{R3(t)\}$ follows a α-stable distribution with a value of α equal to 1.9. 39
4.5 The simulated data was analyzed using proposed methodologies to identify trends, scales, AR models, and distribution tails. These results are presented in rows 1-4. The left column represents the traditional approach, which is designed for signals that follow a Gaussian distribution. The center column represents the robust approach, which is designed for signals that have a non-Gaussian heavy-tailed distribution. The right column is used to evaluate the results using the RMSE for both trend and scale, as well as boxplots for the random components. The random component $\{R3(t)\}$ follows a α-stable distribution with a value of α equal to 1.85. 40
4.6 FEMTO data modelling - the robust approach. Results of each step of the proposed framework: (a) identified trend $T(t)$, (b) random component $\{R(t)\}$, (c) identified scale $SC(t)$, (d) normalized random component $\{R_2(t)\}$, (e) autocorrelation of normalized random component, (f) residuals of AR model, (g) autocorrelation of the residual series, (h) comparison of fitted tails, (i) comparison of fitted distributions PDF. 42

4.7 Wind turbine data modelling - the robust approach. Results of each step of the proposed framework: (a) identified trend $T(t)$, (b) random component $\{R(t)\}$, (c) identified scale $SC(t)$, (d) normalized random component $\{R_2(t)\}$, (e) autocorrelation of normalized random component, (f) residuals of AR model, (g) autocorrelation of the residual series, (h) comparison of fitted tails, (i) comparison of fitted distributions PDF. 43
4.8 The validation of the approach for real datasets (represented by grey lines) is demonstrated: the top panel showcases the FEMTO data, while the bottom panel displays the wind turbine data. Moreover, the constructed quantile lines (depicted by black lines) at the 5% and 95% levels are established based on 100 simulated trajectories corresponding to the fitted non-Gaussian models. 45
4.9 The validation of the approach for real datasets (depicted by grey lines) is presented: the top panel represents the FEMTO data, while the bottom panel displays the wind turbine data. Additionally, the constructed quantile lines (illustrated by blue lines) at the 5% and 95% levels are based on 100 simulated trajectories corresponding to the fitted Gaussian models. 45
4.10 The quantile lines at the 95% level for the FEMTO data (top panel) and wind turbine data (bottom panel) are compared for two models: the suggested non-Gaussian non-homogeneous model and the Gaussian-based model, which includes Gaussian noise and a deterministic trend. The red lines represent the differences between the computed quantile lines. The quantile lines were produced using 100 trajectories of the simulated models in all situations. 47

5.1 Simulated HI: (a) deterministic part with marked changing points, (b) Gaussian noise, (c) non-Gaussian noise with $\nu = 3$, (d) HI in presence of Gaussian noise, (e) HI in presence of non-Gaussian noise. 54
5.2 Changing points detection in the presence of Gaussian noise. 55
5.3 Monte Carlo analysis for changing point detection in presence of Gaussian noise, left subplot is boxplot of detected CP1 and right subplot is boxplot of detected CP2. 55
5.4 MSE for Monte Carlo analysis for changing point detection in presence of Gaussian noise. 56
5.5 Segmentation results for non-Gaussian simulation case with $\nu = 2.1$. . 57
5.6 Monte Carlo analysis for changing point detection in the presence of different levels of non-Gaussian noise, (a) $\nu = 2.1$, (b) $\nu = 3$, (c) $\nu = 5$, (d) $\nu = 10$. 57
5.7 MSE for Monte Carlo analysis for changing point detection in the presence of different levels of non-Gaussian noise, (a) bar plot of MSE for CP1, (b) bar plot of MSE for CP2. 58
5.8 Changing points detection IMS datasets case number 2, (a) result of changing points detection method, (b) envelope spectrum of the raw vibration signal. 59
5.9 Result of changing points detection methods for the FEMTO datasets. 60
5.10 Changing points detection of wind turbine datasets. 61
5.11 Changing points detection wind turbine, (a) raw vibration data, (b) envelope spectrum map. 61

5.12 Percentage error for real datasets changing points detection, (a) percentage error for CP1, (b) percentage error for CP2. 62
5.13 State estimation in the presence of non-Gaussian noise, (a) measurement signal and real state, (b) results of state estimation with KF and MCKF. 65
5.14 Dynamic Bayesian network. 66
5.15 Detection of the stages in the presence of Gaussian noise, (a) HI, (b) probability of stages performed by SKF, (c) probability of stages performed by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF. 70
5.16 Detection of the stages in the presence of non-Gaussian noise, (a) HI, (b) probability of stages by SKF, (c) probability of stages by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF. 71
5.17 Sensitivity analysis, (a) health stage, (b) degradation stage, (c) critical stage. 72
5.18 Specificity analysis, (a) health stage, (b) degradation stage, (c) critical stage. 72
5.19 Segmentation of the FEMTO datasets, (a) HI, (b) probability of stage by SKF, (c) probability of stage by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF. 73
5.20 Segmentation of the wind turbine datasets, (a) HI, (b) probability of stage by SKF, (c) probability of stage by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF. 74

6.1 Complete procedure for estimation of the distribution of RUL. 78
6.2 Concept of RUL of the equipment. 79
6.3 Final results of predicted RUL in the presence of the Gaussian noise, top panel: HI and EOL, bottom panel: predicting RUL results performed by EKF, UKF, and proposed MCEKF, and the real RUL $\pm$20% accuracy bound. 83
6.4 PDF of the estimated RUL in the presence of the Gaussian noise, changing over time. 83
6.5 The 90% CI of RUL estimated using MCEKF for the simulated HI in the presence of Gaussian noise. 83
6.6 Simulated HI in the presence of non-Gaussian noise. 84
6.7 Final results of predicted RUL in the presence of the non-Gaussian noise, top panel: HI and EOL, bottom panel: the mean of the predicted RUL performed by EKF, UKF, and proposed MCEKF, and the real RUL $\pm$20% accuracy bound. 85
6.8 PDF of the estimated RUL in the presence of the non-Gaussian noise, changing over time. 85
6.9 The 90% CI of the RUL estimated by using MCEKF for simulated HI in the presence of non-Gaussian noise. 85
6.10 MAE based on 100 simulated cases in the presence of the Gaussian noise. 86
6.11 Boxplots of MAE based on 100 simulated cases in the presence of the non-Gaussian noise (Student's t distributed noise with $\nu = 2.1$). 86

6.12 Boxplots of MAE for 100 simulated trajectories for case of non-Gaussian noise for different levels of non-Gaussianity (Student's t distributed noise with different levels of the ν value). left panel is represented boxplot of MAE for provided result by MCEKF, middle panel is represented boxplot of MAE for provided result by EKF, and right panel is represented boxplot of MAE for provided result by UKF. 87
6.13 Boxplot of MAE between real RUL and mean of estimated RUL for 100 simulated trajectories for case of Gaussian noise with different kernel size σ for MCEKF. 87
6.14 Boxplot of MAE between real RUL and mean of estimated RUL based for 100 simulated trajectories for case of non-Gaussian noise (Student's t distributed noise with $\nu = 2.1$) with different kernel size σ for MCEKF. 87
6.15 Final results of estimated RUL for FEMTO dataset, top panel: HI and EOL, bottom panel: predicted RUL results performed by EKF, UKF and proposed MCEKF, and real RUL $\pm$20% accuracy bound. 88
6.16 PDF of estimated RUL, changing over time, obtained by using MCEKF for FEMTO dataset. 89
6.17 The 90% CI of estimated RUL by using MCEKF for FEMTO dataset. . 89
6.18 Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound. 89
6.19 MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL. 89
6.20 (a) Health index (RMS) of IMS dataset: subset 3, bearing 3 (b) FPT detected by SMCKF. 90
6.21 Final results of estimated RUL for IMS dataset, top panel: HI and EOL, bottom panel: predicted RUL results performed by EKF, UKF and proposed MCEKF, and real RUL $\pm$20% accuracy bound. 90
6.22 PDF of estimated RUL, changing over time, obtained by using MCEKF for IMS dataset. 91
6.23 The 90% CI of estimated RUL by using MCEKF for IMS dataset. 91
6.24 Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound. 91
6.25 MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL. 91
6.26 Final results of estimated RUL for wind turbine dataset, top panel: HI and EOL, bottom panel: predicted RUL results performed by EKF, UKF and proposed MCEKF, and real RUL $\pm$20% accuracy bound. . . 92
6.27 PDF of estimated RUL, changing over time, obtained by using MCEKF for wind turbine dataset. 93
6.28 The 90% CI of estimated RUL by using MCEKF for wind turbine dataset. 93
6.29 Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound. 93
6.30 MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL. 93
6.31 Calculated MAE for all datasets. 94

List of Tables

2.1 The contribution of authors of mentioned papers (the numbers refers to the research activities as mentioned). 8

3.1 Main characteristics of the data for three regimes indicated in Fig. 3.11. 27

4.1 The summary of the framework for data modelling - classical and robust approaches. 35
4.2 Identified characteristics for the FEMTO data set. 42
4.3 Identified characteristics for the wind turbine dataset. 44

6.1 Selected parameters value for MCEKF in case of FEMTO dataset. . . . 89
6.2 Selected parameters value for MCEKF in case of IMS dataset. 92
6.3 Selected parameters value for MCEKF in case of wind turbine dataset. 94

List of Abbreviations

PHM	Prognostics and health management
CM	Condition monitoring
RUL	Remaining useful life
KF	Kalman Filter
EKF	Extended Kalman Filter
ML	Machine learning
QKF	Quadratic Kalman filter
UKF	Unscented Kalman filter
CKF	Cubature Kalman filter
SMC	Sequential Monte Carlo
PF	Particle filter
MSE	Mean square error
MCEKF	Maximum correntropy criterion switching Kalman filter
EOL	End of the life
FPT	First prediction time
HI	Health index
PDF	Probability density function
CI	Confidence interval
MAE	Mean absolute error
IMS	Intelligent maintenance system data set
PHM	Prognostics and health management
RMS	Root mean square
TRB	Tapered roller bearing
SCADA	Supervisory control and data acquisition
CMS	Condition monitoring systems
CM	Corrective maintenance
PM	Preventive maintenance
TBM	Time-based maintenance
OEM	Original equipment manufacturer
OSA-CBM	Open system architecture for condition-based maintenance
ANNs	Artificial neural networks
NF	Neuro-fuzzy
SVM	Support vector machine
GPR	Gaussian process regression
KNN	K-nearest neighbour
FFNNs	Feed forward neural networks
PDF	Probability density function
AR	Auto-regressive
BM	Brownian motion
FBM	Fractional Brownian motion
FLSM	Fractional Leavy stable motion
GC	Generalized Cauchy

CP	Changing point
CI	Confident interval
FPT	First prediction time
OLS	Ordinary least square
IRLS	attentively reweighed least square
ST	Student's t distribution
MLE	Maximum likelihood estimation
IQR	Interquartile range
HMMR	Hidden Markov model regression
HLP	Hidden logistic process
PHM	Prognosis and health managment

List of Symbols

S	Synthetic degradation data
xs	Dynamic state
SM	Switching variable
R	Random component
T	Deterministic component of signal
t	Time
p	Order of autoregressive time series
SC	Scale of random component
a	Constant coefficient
b	Constant coefficient
c	Constant coefficient
X	Random variable
x	Random sample
F_X	Cumulative distribution function
KS	Kolmogorov- Smirnov (KS) statistic
$\mathbf{x}$	Sample mean for the signal
h	Lag
u	Observation vector
v	Observation vector
f	Polynomial function
K	Number of distinct degradation stages
Z	Hidden states
y	Observation
m	Observation noise
q	Process noise
H	Observation matrix
A	State-transition model
W	Weights matrix
P	Covariance matrix
xs	Actual state
τ_1	Time of the changing point from first to second regime
τ_2	Time of the changing point from second to third regime
ϕ	coefficient of autoregressive time series
σ	Standard division
$\hat{\rho}_X$	Efficient estimator of the ACF
$\hat{\gamma}_X$	Sample autocovariance function (ACVF)
θ	Constant coefficient
ϕ_{Tu}	Tukey biweight function
ν	Degree of freedom Student's t distribution
p_ν	Probability density function (PDF) of Student's t distribution
β	Polynomial coefficient
ϵ	Gaussian white noise

$\overline{K}$	Kalman gain
y	Observation

Chapter 1

Introduction

The advancement of technology in data acquisition, transmission, and storage has facilitated the emergence of long-term historical data. Within the industrial sector, there is notable growth in the production of data at different rates and scales, originating from diverse sources such as sensor-rich condition monitoring systems (CMS) and supervisory control and data acquisition (SCADA) systems. These data sources constitute a rapidly growing repository for research in CBM. This is particularly noteworthy, as both researchers and experts recognize the massive potential of extracting meaning full information from these term historical data.

1.1 Evolvement of maintenance strategy

A significant change in maintenance strategies has taken place within the domain of maintenance research, as illustrated in Fig.1.1. In essence, a transition has occurred from corrective maintenance to preventive maintenance (PM). Throughout history, corrective maintenance practices have been characterized by a "fail and fix" mentality, in which maintenance actions were performed in response to apparatus failure and were deemed reactive in nature. However, this reactive approach frequently resulted in unanticipated shutdowns, which posed environmental and safety risks and caused significant economic losses.

Companies implemented a PM strategy, which involves proactively performing maintenance and repair activities prior to asset failure, in order to mitigate the potential repercussions. PM may be determined by the equipment's health condition (referred to as CBM) or by a predetermined schedule (known as time-based maintenance). At the outset, project management endeavors were frequently executed at predetermined time intervals, depending on the knowledge and proficiency of engineers and technicians or the guidance provided by the original equipment manufacturer (OEM) through laboratory experiments and reliability theory.

In order to recommend PM intervals, computational simulations of complex systems have been implemented in tandem with developments in computing technology. However, there are two significant disadvantages to a predetermined maintenance strategy, despite its ability to reduce unanticipated breakdowns. In the first place, excessive maintenance may result in substantial costs, which do not necessarily enhance the dependability of the apparatus. Furthermore, this approach operates under the assumption that failure behavior is predictable, which may not accurately represent the complex interplay between systems and components, environmental influences, equipment aging, and other variables that frequently affect failure behavior.

In contrast, predictive maintenance was suggested as a remedy for these issues. CBM initiates maintenance tasks when necessary in accordance with the health condition of the apparatus to forecast future failures. CBM, in contrast to predetermined

maintenance, modifies maintenance activities in response to condition data, thereby documenting the equipment's ever-changing state. This process involves the collection and measurement of extensive volumes of data via condition monitoring, in situ inspection, or testing. Subsequently, a multitude of data analysis methodologies, including signal processing, machine learning, and data mining, are implemented to evaluate the health of the apparatus and aid in the formulation of maintenance strategies.

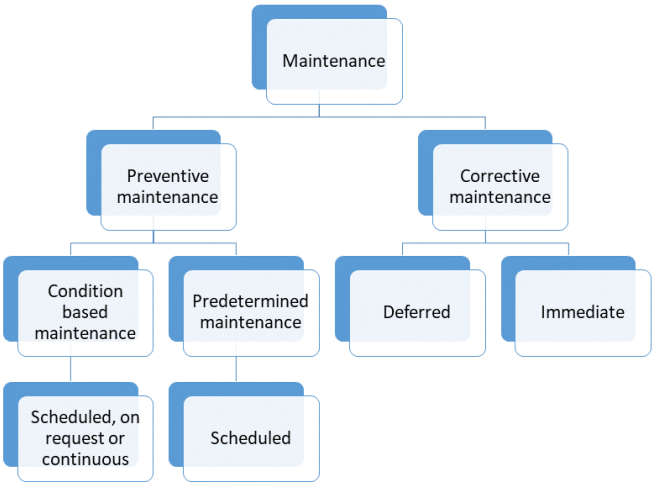


FIGURE 1.1: Maintenance strategy.

1.2 Condition based maintenance

CBM is a form of preventive maintenance that combines condition monitoring, inspection, testing, analysis, and subsequent maintenance actions [1]. The core components of CBM implementation include continuous, periodic, or on-demand condition monitoring, inspection, and testing, tailored to the criticality of the monitored item. Subsequent analysis evaluates the health condition and predicts the RUL of the item, forming the foundation of a CBM strategy. The final step involves determining maintenance actions through a decision-making process that considers maintenance resources, operational contexts, and inputs from other systems.

An open system architecture for condition-based maintenance (OSA-CBM) has been devised in accordance with the functional specifications outlined in ISO-13374 for condition monitoring and machinery diagnostics [2]. As illustrated in Fig. 1.2, the OSA-CBM architecture comprises seven layers: data acquisition, data preprocessing and health index construction, state detection, machine health assessment, prognostics, decision support, and presentation, which will be discussed in the next step.

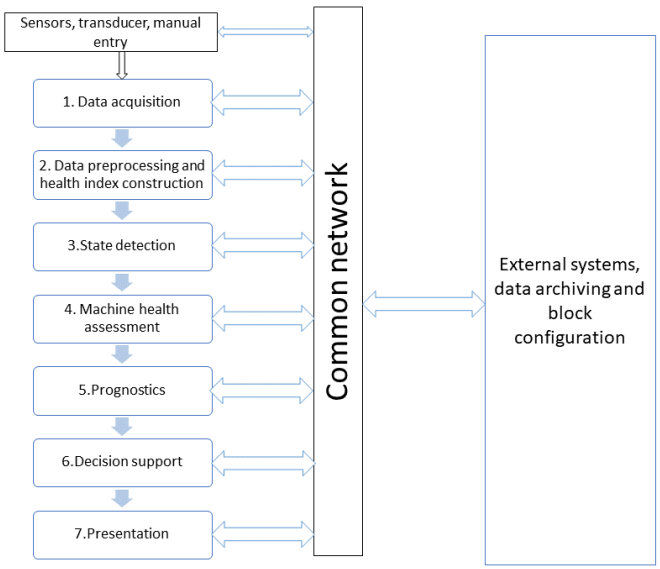


FIGURE 1.2: Condition based maintenance workflow.

- **Data acquisition (Step 1)**: The initial stage involves collecting raw data from various sources, such as sensor measurements, thermal images, acoustic recordings, and machinery vibrations. These data streams, originating from various systems with differing sampling rates, serve to monitor machinery operation, presenting integration challenges within the maintenance domain.
- **Data preprocessing and health index construction (Step 2)**: This phase, similar to data preparation in typical data mining processes, involves the transformation of raw data into a final dataset for analysis. Techniques such as data cleansing, feature selection, and standardization ensure an accurate representation of machinery condition, necessitating tailored approaches.
- **State detection (Step 3)**: Also termed fault detection, this step compares data with predefined limits to determine proper functioning or anomalies. Sensitivity to changes in the operational context and adaptability to new environments are crucial.
- **Machine health assessment (Step 4)**: This phase evaluates degradation, diagnosing faults with associated confidence levels, considering the health history and operational context.
- **Prognostics (Step 5)**: Forecasting future states involves estimating RUL using prognostic and predictive models, with the assignment of confidence levels reflecting inherent uncertainty.
- **Decision support (Step 6)**: Recommendations based on predictions, mission profiles, objectives, and resource limitations require detailed advisories for maintenance scheduling.
- **Presentation (Step 7)**: An interactive interface centralizes the data and the results, enabling deeper analysis.

The OSA-CBM architecture offers a comprehensive framework for CBM, with each layer requiring individualized treatment and the development of specific techniques. Ideally, the tasks delineated in these layers should be executed sequentially

to autonomously schedule CBM activities. However, in certain instances, the seamless progression through this linked chain may be hindered due to a lack of knowledge in particular layers. For example, the absence of suitable prognostic models can impede the automatic execution of the prognostic task. In such scenarios, the use of expert knowledge and experience becomes imperative to perform subsequent procedures. However, the preceding procedures remain informative and furnish a robust factual foundation for human judgments. The focal point of this thesis lies in investigating tasks spanning from layer 4 to layer 5 within the OSA-CBM architecture.

1.3 Machinery's long term condition monitoring data

Condition monitoring data, sourced from various channels, contain valuable information about the systems. In particular, in the domain of CBM, data collected from different machines play a pivotal role. This data, acquired over hours, days, months, or even years, is used to monitor the health of machines and encompasses a wealth of information about the machine's historical operating conditions. Such data have immense potential for CBM applications. Figs. 1.3, and 1.4, illustrate some raw historical degradation data collected from the bearing.

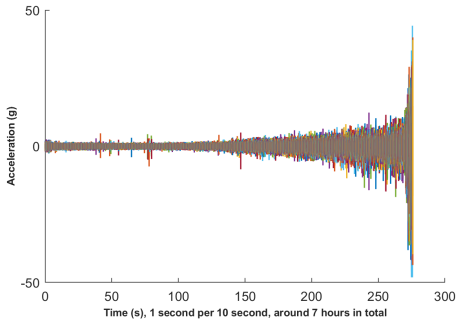


FIGURE 1.3: Recorded vertical vibration of degraded bearing over time.

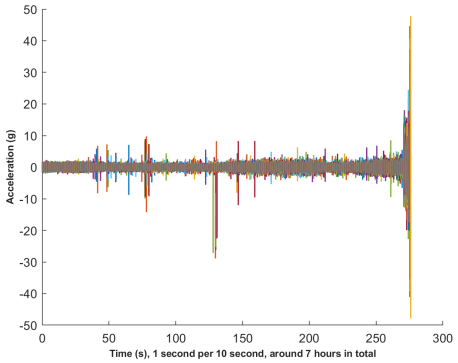


FIGURE 1.4: Recorded horizontal vibration of degraded bearing over time.

However, as discussed, working with raw signals may not be optimal, and it may be better to extract features from the raw signal and use them to represent the health state of the machine. This approach offers several benefits. Firstly, it reduces the volume of the signal for further processing. Additionally, extracting features from the raw signal allows for the derivation of knowledge-based features that may better represent the health state of the machine compared to the raw signal alone. Moreover, feature extraction provides an opportunity to view the signal in different ways, potentially revealing insights that may not be apparent when analyzing the raw signal directly.

Chapter 2

Problem formulation and state of the art

This chapter focuses on investigating the research area, which includes problem formulation, purpose and objectives, research questions, limitations, and state-of-the-art.

2.1 Problem formulation

Machine health assessment and prognosis are two crucial phases of CBM. As discussed in Chapter 1, the input for these phases is the health index, which reflects the health state of the system. However, in many industrial settings, machines operate under varying and harsh conditions, leading to non-stationary and non-linear behavior in the health index. Additionally, evidence of non-Gaussian noise, particularly heavy-tailed noise, can often be observed in the health index.

Various sources can contribute to this non-Gaussian noise, for example, in mining applications, falling ore onto equipment can generate such noise. Similarly, in aviation and wind turbines, see Fig. 2.1, turbulence can act as another source of non-Gaussian noise. Additionally, unpredictable electromagnetic behavior can also generate non-Gaussian noise.

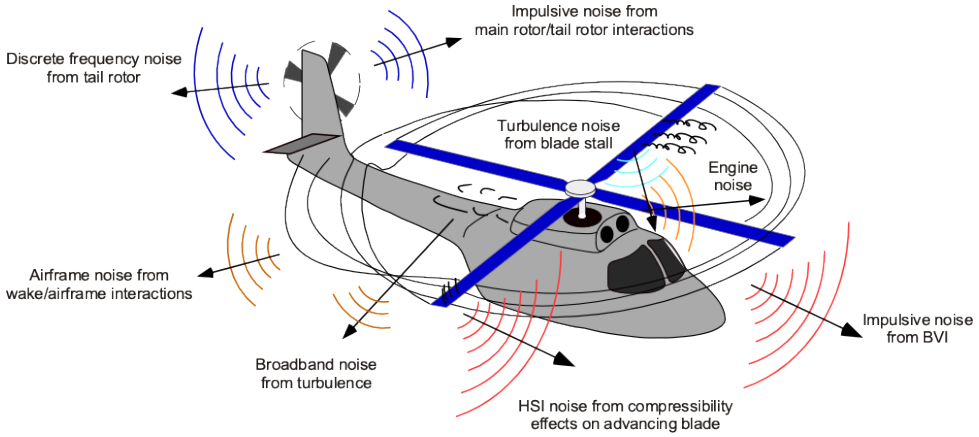


FIGURE 2.1: Helicopter noise source [3].

Fig 2.2 illustrates a health index from a wind turbine (further details are provided in Chapter 3), serving as an example of non-stationary, non-linear behavior with evidence of non-Gaussian noise in real-world applications.

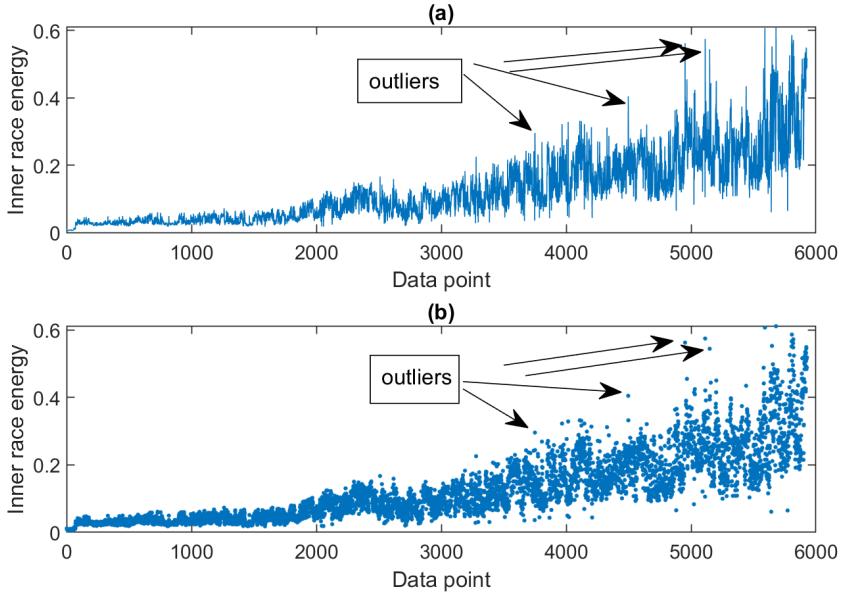


FIGURE 2.2: Wind turbine health index.

As depicted in Fig. 2.2, numerous outliers are evident over time. The presence of these outliers can significantly impact the performance of classical approaches, which typically assume Gaussian noise, when attempting to characterize the health index. Moreover, modeling such sequences of health indices while accounting for non-Gaussian noise poses a considerable challenge.

Facilitating maintenance decision making and developing effective diagnostics and prognostic models in harsh environments are crucial research directions. Therefore, this study places emphasis on addressing the challenges associated with non-Gaussian noise and non-stationary behavior in long-term health index data for machine health assessment and prognostics applications.

2.2 Purpose and objectives

This section outlines the purpose and objectives of the research.

The primary objective of the research is to investigate, explore, and develop approaches to the assessment and prognosis of machinery health based on long-term data.

The specific objectives include:

- Investigation of the effect of non-Gaussian noise on modeling long-term health index data.
- Development of robust framework for identifying and modeling long-term health index data in the presence of non-Gaussian noise.
- Development of segmentation approach based on long-term data for the purpose of machine health assessment while considering non-Gaussian noise.
- Development of robust probabilistic models for machine prognostics based on long-term health index data in the presence of non-Gaussian noise.

2.3 Research questions

To accomplish the stated objectives, the following research questions have been formulated:

- How can a model be developed to effectively handle long-term historical degradation data in the presence of non-Gaussian noise?
- What methods should be employed to properly identify, characterize, and extract useful information from long-term historical degradation data for the development of robust data-driven degradation models?
- How can the segmentation of long-term degradation data be robustly performed, particularly in the presence of non-Gaussian noise?
- What approaches should be employed to robustly predict RUL based on long-term degradation data, especially in the presence of non-Gaussian noise?

2.4 Author's Contribution to the Field

In this thesis, we explore the modeling, analysis, segmentation, and prediction of RUL using long-term HI data. Our primary focus is on HI data with non-Gaussian (heavy-tailed) noise. While the use of long-term HI data in CBM programs is becoming more popular, there is still a lack of well-established references for handling HI data with non-Gaussian noise. Therefore, this thesis proposes a new framework for modeling and identifying long-term HI data in the presence of non-Gaussian noise (see Chapter 4).

Additionally, this dissertation introduces robust approaches for segmenting HI data into three stages based on historical data and an online method for real-time segmentation into three stages (see Chapter 5). Finally, it presents a probabilistic method for predicting RUL using an exponential state degradation model and a robust filter (see Chapter 6). The main goal is to contribute to the field by establishing statistical-based modeling approaches.

We have published several articles related to long-term HI data modeling, identification, segmentation, and RUL prediction.

The results obtained in this thesis were published in the following articles:

I Żuławiński, W., Maraj-Zygmąt, K., **Shiri, H.**, Wyłomańska, A., and Zimroz, R. (2023). Framework for stochastic modelling of long-term non-homogeneous data with non-Gaussian characteristics for machine condition prognosis. Mechanical Systems and Signal Processing, 184, 109677.

II **Shiri, H.**, Zimroz, P., Wodecki, J., Wyłomańska, A., Zimroz, R., and Szabat, K. (2023). Using long-term condition monitoring data with non-Gaussian noise for online diagnostics. Mechanical Systems and Signal Processing, 200, 110472.

III **Shiri, H.**, Zimroz, P., Wodecki, J., Wyłomańska, A., and Zimroz, R. (2023). Data-driven segmentation of long-term condition monitoring data in the presence of heavy-tailed distributed noise with finite-variance. Mechanical Systems and Signal Processing, 205, 110833.

IV **Shiri, H.**, Zimroz, P., Wyłomańska, A., and Zimroz, R. (2024). Estimation of machinery's remaining useful life in the presence of non-Gaussian noise by using a robust extended Kalman filter. Measurement, 114882.

In paper I, a robust framework for modeling and identifying long-term HI data in the presence of non-Gaussian noise was introduced. In paper II, a robust segmentation approach based on dynamic trends and historical HI data to segment HI into three distinct stages was proposed. Additionally, in paper III, an online robust switching method for segmenting HI data into three stages was introduced. Finally, in paper IV, a robust probabilistic approach based on Bayesian theory to predict RUL was proposed.

2.5 Authorship of mentioned papers

The Table 2.1 shows the participation of each author in the listed papers with respect to the following activities.
1-Developing the core concepts of the issue;
2-Conducting the research;
3-Drafting the paper;
4-Reviewing significant intellectual materials;
5-Giving the final approval for submission.

	Paper I	**Paper II**	**Paper III**	**Paper IV**
Hamid Shiri	2-5	1-5	1-5	1-5
Pawel Zimroz	-	2,4,5	2,4,5	2,4,5
Radoslaw Zimroz	1,4,5	1,4,5	1,4,5	1,4,5
Agnieszka Wyłomańska	1,4,5	1,4,5	1,4,5	1,4,5
Wojciech Żuławiński	1-5	-	-	-
Katarzyna Maraj-Zygmąt	1-5	-	-	-
Jacek Wodecki	-	4-5	4-5	-
Krzysztof Szabat	-	4-5	-	-

TABLE 2.1: The contribution of authors of mentioned papers (the numbers refers to the research activities as mentioned).

2.6 Scope of research

The scope of this research encompasses the study of knowledge-based data-driven approaches for machine health assessment and prognostics, focusing on long-term health index data in the presence of non-Gaussian noise. Specifically, the research aims to:

1. Develop a robust framework for identifying and characterizing long-term health index data for modeling purposes in the presence of non-Gaussian noise, see Chapter 4.

2. Develop offline and online segmentation models for long-term health index data without relying on threshold values for machine health assessment in the presence of non-Gaussian noise, see Chapter 5.

3. Develop a probabilistic approach for machine prognostics based on health index data in the presence of non-Gaussian noise, see Chapter 6.

These research objectives also involve comparing the proposed approaches with classical methods based on Gaussian assumptions to elucidate the impact of non-Gaussian noise on performance results. The performance of all approaches is independently verified using synthetic data and real datasets.

2.7 Limitations of this research

The limitations of this thesis can be described as follows:

1. In this thesis, the effect of health index construction and the methodology for extracting this health index are not investigated. This aspect could have significant implications for the performance and accuracy of the developed approaches. In this thesis, it is assumed that the selected health index is the best indicator for describing the degradation process.

2. The non-linear behavior of the health index is not thoroughly investigated in detail. Non-linear in the health index can introduce complexities that were not fully addressed in this investigation.

3. In Chapters 5 and 6, approaches were developed based on parameters identified and tuned using historical data. However, in some cases, historical data may not be available or may not adequately represent future conditions, potentially limiting the generalizability of the developed approaches.

4. To estimate the RUL, the results provided in this thesis are based on known end-of-life (EOL) thresholds. However, determining the exact EOL threshold can be a challenging task that requires further study beyond the scope of this thesis

2.8 State of the art

This section will cover state of the art of modeling, segmentation and prediction of the long-term condition monitoring data.

2.8.1 Modeling of long-term health index data

Recent years have witnessed significant research endeavors in degradation modeling and prediction, broadly delineated into three categories, see Fig. 2.3: data-driven models [4, 5, 6, 7], physics-based models [8], and hybrid models [9, 10], see Fig.2.3.

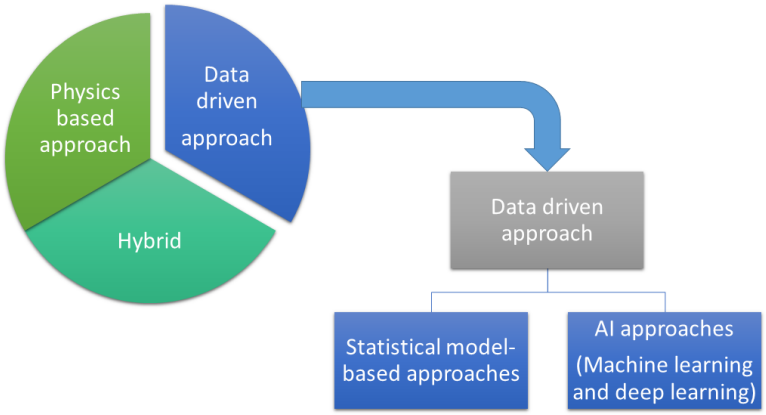


FIGURE 2.3: Long-term degradation modeling, prediction approaches.

Physics-based models aim to accurately represent the degradation process by utilizing mathematical formulations rooted in the principles of failure and damage mechanics. These models are capable of producing highly precise results, as they are grounded in well-established physical laws and mechanisms. The parameters of physics-based models are intrinsically linked to material properties and stress levels, which are typically determined through specific experimental methods, finite element analysis, or other appropriate techniques. One of the most widely utilized physics-based models for predicting the RUL of machinery is the Paris-Erdogan model. Initially proposed in [11] to describe crack growth behavior, the Paris-Erdogan model has since undergone numerous adaptations and applications within machinery prognostics [12, 13].

For example, Wang et al. [14] and Lei et al. [15] modified the Paris-Erdogan model to develop empirical models for machinery RUL prediction. Liao [16] and Sun et al. [17] further advanced the Paris-Erdogan model by converting it into a state-space model, enhancing its predictive capabilities. Beyond the Paris-Erdogan model and its derivatives, several other physics-based models have been employed in the field of machinery RUL prediction. Oppenheimer et al. [18] utilized the Forman crack growth law [19] to estimate the RUL of the cracked rotor shafts. Similarly, Baraldi et al. [20], and Hu et al. [21] applied the Norton law to describe creep behavior in turbines, combining this approach with Kalman filtering (KF) and particle filtering techniques for RUL prediction. Furthermore, Chan et al. [22] developed a time-dependent crack growth model specifically for turbo-propulsion systems, while El-Tawil et al. [23] incorporated stochastic descriptions into a non-linear damage law to predict the RUL of pipeline tubes.

However, the accuracy of these models depends on the availability of detailed and precise physical representations of the system being studied. In many cases, such detailed information may be difficult to obtain or may require extensive data collection and analysis. Additionally, the implementation of physics-based models often involves significant computational resources and time, due to the complexity of the underlying equations and the need for high-resolution simulations. Consequently, while physics-based models are powerful tools for understanding and predicting degradation processes, their practical application may be limited by these

demanding requirements.

Data-driven approaches focus on constructing models that characterize degradation processes using historical data. These methods are particularly useful in scenarios where developing physics-based models is challenging. Data-driven approaches can be categorized into two main subclasses: AI-based methods (e.g., [24, 25]) and statistical based-model methods (e.g., [26, 27, 28]).

AI-based methods are highly effective in modeling and predicting degradation processes. AI methods use sophisticated artificial intelligence algorithms to discover machinery degradation trends from observations rather than physics-based or statistical models. These methods are particularly adept at addressing prognostic challenges in complex mechanical systems where degradation processes are difficult to model using traditional methods. AI-based techniques are the second largest category of articles, behind statistical models-based approaches [29]. Although popular, AI approaches are sometimes criticized for their lack of transparency, garnering the nickname "black boxes." Because AI model decision-making processes are hard to understand, this opacity exists.

In machinery prognostics, artificial neural networks (ANNs), neuro-fuzzy (NF) systems, suport vector machine (SVM), K-nearest neighbor (KNN) and Gaussian process regression (GPR) are employed. These methods have distinct benefits and have advanced the prognostics of machinery. Sbarufatti et al. [30] integrated feed-forward neural networks (FFNNs) with sequential Monte Carlo sampling to predict the RUL of fatigue cracks. This hybrid approach leverages the pattern recognition capabilities of FFNNs and the probabilistic framework of Monte Carlo sampling to improve predictive accuracy. Similarly, Pan et al. [31] and Xiao et al. [32] employed FFNNs for multistep ahead predictions of bearing health states, demonstrating the efficacy of neural networks in forecasting the progression of bearing degradation. These studies highlight the potential of combining FFNNs with advanced sampling techniques to improve prognostic performance in various applications. Recurrent neural networks (RNNs) are extensively utilized in RUL prediction due to their proficiency in handling explicit time-series data. Zemouri et al. [33] introduced a recurrent radial basis function network and applied it to predict the RUL of machinery, leveraging the network's capability to model temporal dependencies. Furthermore, Malhi et al. [34] proposed a competitive learning-based approach to enhance the training methodology of RNNs, with the objective of improving the accuracy of long-term prediction. These advances underscore the effectiveness of RNNs and their variants in accurately forecasting machine degradation over extended periods. El-Koujok et al. [35] presented a NF prognostic technique that balances model complexity and accuracy using parsimony. This ensures the efficiency of the model and the predictive power of the model. Ishibashi et al. [36] created a fuzzy rule-based genetic prognostic system with genetically optimized membership functions. This strategy fine-tunes fuzzy rules using evolutionary methods to improve system flexibility and accuracy. Liu et al. [37] used two covariance functions to improve the long-term prediction of GPR. This method captures global deterioration patterns and local regeneration behaviors, improving predictions. Aye et al. [38] predicted the degradation of the bearing of the rolling element using an integrated GPR model. Their research shows that GPR can describe complicated deterioration patterns and provide an accurate prognosis.

However, they typically require large amounts of training data to achieve accurate and reliable results. This dependency on extensive datasets can pose significant

limitations in real-world applications, where such comprehensive data may not be readily available.

Statistical based-model techniques, also known as empirical model-based approaches, estimate machinery's RUL using empirical models. Methods often use a probability density function (PDF) to forecast RUL based on data [6]. Without applying physical principles, these strategies fit observational data into random coefficient or stochastic process models using probabilistic methods to create RUL prediction models.
For degrading uncertainty, random variances are added to the model parameters. Variances originate from temporal, unit-to-unit, and measurement sources [39]. Statistical based-model techniques are especially successful in characterizing deterioration uncertainty and its impact on RUL forecasts. Statistical based-model help machinery prognostics by using these uncertainties to understand and forecast equipment degradation.

One of the most commonly used statistical-based models is the autoregressive (AR) model, which implies that a machine's future state value is linearly related to previous observations and random errors [40]. Qian et al. [41] predicted bearing degradation using the AR model, which captured linear relationships in the degradation data. Barraza-Barraza et al. [42] used three AR models with exogenous factors to predict the RUL of aluminum plates with fatigue fractures, demonstrating the flexibility of the model. Escobet et al. [43] predicted the RUL of a conveyor belt system using an AR model, demonstrating the adaptability of AR models in prognostic applications. Statistical based-model techniques have been thoroughly reviewed and categorized into different groups [6, 7]. To avoid repetition, this thesis categorizes these techniques into two subgroups, as illustrated in Fig. 2.4. The first main group consists of methods developed based on Gaussian assumptions, while the second group encompasses methods based on non-Gaussian assumptions.

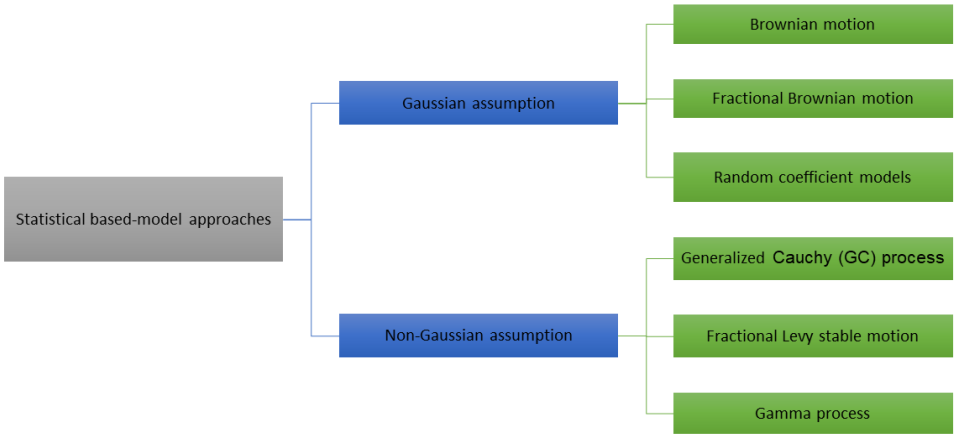


FIGURE 2.4: Statistical based-model approaches according the assumption of the noise.

Wiener process models usually include drift and diffusion terms following Brownian motion. Bian et al. [44] developed a covariance-dependent degradation model

to better capture variability. Doksum et al. [45] used Wiener process models to predict RUL in accelerated degradation testing with variable stress. Whitmore et al. [46] suggested a time scale transformation technique to stabilize Wiener process models by converting time-varying degrading drift to constant drift. Tseng et al. [47] used Wiener process models to predict RUL and plan appropriate burn-in practices. Park et al. [48] created Wiener process models that use a cumulative damage method to better understand the accumulation of damage over time. Gebraeel et al. [49] suggested a Brownian motion error-based exponential model for improved prediction accuracy.
BM models assume the Markov property, which states that the future state relies only on the present state and not on the previous behavior. This assumption does not always hold in practice. Xi et al. [50] developed a degradation model using fractional Brownian motion (FBM) to estimate RUL, addressing long-range dependencies in degradation processes.
BM and FBM have stationary increments. The increments of both systems have the same distribution. However, FBM may represent non-stationary degradation processes when only the drift factor is non-linear. The non-stationary degradation model is taken solely into account in increment expectations, since drift terms represent the deterministic element of the process [51]. The variety of degradation processes is usually predictable and random. Non-stationary properties of a random portion must be considered.

Most random-coefficient models use normally distributed random coefficients to represent the stochasticity of degradation. Lu et al. [52] and Meeker et al. [53] used Monte Carlo simulation to forecast the PDF of RUL for machinery degradation processes using non-linear mixed-effects models. Bayesian parameter estimation was used by Gebraeel et al. [49] to suggest an exponential model with random error terms. Park et al. [54] defined degradation processes using a non-linear random-coefficient model and estimated parameters using maximum likelihood estimation. This method captures complicated degradation patterns and estimates parameters for a better prognosis. Coble et al. [55] developed a prognostic approach that incorporates previous beliefs into random coefficient models, allowing the prediction of the RUL for machining tools. This novel method improves the strength and precision of prognostic predictions by integrating past information into the modeling procedure. Random coefficient models have the benefit of using variations in random coefficients to provide prediction results which offer a PDF of RUL. However, the assumption that the random coefficients follow a Gaussian distribution may restrict their practicality. Furthermore, these models lack the ability to include temporal fluctuations in the RUL predictions, which are essential to correctly represent the dynamic character of degradation processes as they unfold over time.

In the literature, non-Gaussian processes such as the Gamma process, Generalized Cauchy (GC) process, and processes with heavy-tailed distributions, such as fractional Lévy stable motion (FLSM), are also considered. These alternative processes offer valuable alternatives to Gaussian models, allowing for the incorporation of non-normal distributions and heavy-tailed behaviors [56]. By considering non-Gaussian processes, researchers can more accurately capture the complex variability and extreme events often observed in real-world degradation processes.

Gamma process models assume that the increments of degradation processes at

disjoint time intervals are independent random variables with a Gamma distribution. This assumption enables the characterization of degradation processes with non-negative, continuous-time increments, making Gamma process models particularly suitable for modeling phenomena such as degradation, wear, or fatigue where positive accumulative effects occur over time. Noortwijk et al. [57] presented a comprehensive review of the progress made in stochastic degradation modeling using Gamma processes. They summarized many applications and achievements in this subject. Kuniewski et al. [58] hypothesized that degradation process begins at an unpredictable moment, following a non-uniform Poisson process, and then progresses according to a Gamma process. Bagdonavicius et al. [59] analyzed degradation processes using a Gamma process model. They took into account the impact of variables and traumatic events on degradation phenomena. Lawless et al. [60] created a manageable Gamma process model that considers the impact of random degradation rates. This improves the model's ability to accurately anticipate complicated deterioration patterns. Similarly to BM models, Gamma process models are adept at capturing the temporal variability of degradation processes. However, they also have certain limitations. One such limitation is that the noise in Gamma process models must follow a Gamma distribution, restricting their effectiveness to describing monotonic degradation processes. This constraint may limit their applicability in scenarios where degradation exhibits non-monotonic or non-linear behaviors.

The GC process model is a stochastic process used to characterize random variables with heavy-tailed distributions. It is an extension of the Cauchy distribution and offers a flexible framework for modeling phenomena characterized by extreme events and outliers. Liu et al. [61] introduced a novel RUL prediction model based on the GC process, which is characterized by independent parameters. Their model leverages the flexibility and heavy-tailed nature of the GC process to improve the accuracy of RUL predictions for various engineering systems. Similarly, Hong et al. [62] proposed an iterative model of the GC process with long-range dependence (LRD) characteristics to predict the RUL of lithium-ion batteries. By incorporating LRD features into the GC process, their model accounts for the persistence of degradation patterns over time, leading to more reliable and robust predictions of battery RUL. The GC process may lack the mathematical tractability and simplicity of other stochastic processes, such as the Gaussian process. This can make it more difficult to analyze and interpret the results of models based on the GC process, especially in complex engineering systems, where interpretability is crucial. Furthermore, the GC process may require a larger amount of data to accurately estimate its parameters compared to simpler models, which can be a limitation in scenarios where data availability is limited or costly to obtain.

Lévy stable motion is a stochastic process characterized by its stability property, which means that the sum of a large number of independently and identically distributed random variables converges in distribution to a Lévy stable distribution. These distributions have heavy tails making them suitable for modeling phenomena with extreme value. The FLSM is a generalization of the Lévy stable motion that incorporates fractional integration or differentiation into the process. This allows for the modeling of phenomena with long memory or fractal properties, where the correlation between observations decays slowly over time. The FLSM is particularly useful for modeling time series data with long-range dependence. Li et al. [56] developed a novel framework for the prediction of RUL based on performance evaluation and geometric fractional Lévy stable motion (GFLSM) with adaptive non-linear

drift. In their approach, multiple degradation stages are consolidated into a relatively unified model using GFLSM. The long-range dependence and self-similarity of the degradation process are captured by the relationship between the Hurst exponent and the stability exponent. Duan et al. [63] introduced a multi-modal FLSM degradation model to predict the technical life or RUL of the equipment. This model incorporates multiple modes of degradation, allowing the characterization of complex degradation patterns observed in real-world systems. Using FLSM, which captures the long-range dependence and self-similarity of degradation processes, the proposed model offers an effective framework for accurately predicting equipment lifespan and RUL.

Indeed, the mentioned stochastic models, including FLSM, have their strengths and limitations, and their suitability depends on specific assumptions and characteristics of the data. Although FLSM and other heavy-tailed models may better capture the behavior of real-world data compared to Gaussian models, they may still have limitations in certain scenarios.
One limitation is that the scale parameter, which determines the spread or variability of the distribution, is often assumed to follow a specific function, such as a power-law function in the case of FLSM. This assumption may oversimplify the variability in real data, particularly in cases where the data exhibit complex and heterogeneous behavior. Furthermore, the assumption of a specific heavy-tailed distribution, such as the α-stable distribution in the case of FLSM, may not always accurately capture the true distribution of the data. Different types of heavy-tailed distribution may be more appropriate for different datasets, and a one-size-fits-all approach may not be suitable.

Furthermore, the choice of distribution for the random part of the model may vary depending on the characteristics of the data. Although α-stable distributions are commonly used to model heavy-tailed behavior, other distributions, such as the generalized extreme value distribution or the Cauchy distribution, may also be suitable for certain types of data.

Alternative methods for modeling health index data include the Kalman filter or its variants, Markov jump systems, state-space models, and other conventional approaches. Moreover, reliability models have been proposed that clarify HI data fluctuations.

In summary, while stochastic models such as FLSM offer valuable tools for capturing heavy-tailed behavior in real-world data [56], it is essential to consider the assumptions and limitations of these models and to carefully assess their suitability for the specific characteristics of the data at hand. In addition, flexibility in model selection and parameter estimation techniques may be necessary to accommodate the diverse and complex nature of real-world data sets.
In addressing the limitations of individual categories in the prediction of RULs, researchers have explored hybrid approaches that integrate the strengths of different methods. These hybrid methodologies aim to leverage the advantages of diverse approaches to improve predictive precision and robustness. Some notable strategies include:

- **Combining degradation models with particle filters or Kalman filter:** Several studies have utilized a combination of different degradation models, such

as physics-based, data-driven, or statistical models, and integrated them with particle filters or Kalman filter family for RUL prediction [64]. This approach enables the incorporation of both the deterministic and stochastic aspects of degradation processes, leading to improved prognostic capabilities.

- **Fusion strategies for predictive maintenance approaches:** Researchers have developed fusion strategies to combine predictions from multiple approaches, such as physics-based, data-driven, and statistical models, into a unified RUL estimate [65]. These fusion strategies may include techniques from machine learning, statistical inference, or expert systems to integrate predictions and make a final decision.

- **Hybridization of machine learning and statistical models**: Hybrid approaches have been proposed that combine machine learning techniques, such as artificial neural networks (ANNs), with statistical models, such as AR models [66]. By integrating the complementary strengths of these methods, hybrid approaches can capture complex patterns in data while incorporating uncertainty quantification from statistical models.

- **Integrated diagnosis and prognosis frameworks**: Some researchers have developed integrated frameworks that combine diagnosis and prognosis functions to provide a comprehensive solution to predict RUL [67]. These frameworks leverage both historical data and real-time monitoring information to continuously update and refine RUL estimates, enhancing reliability and timeliness in prognostic assessments.

For further information on hybrid RUL prediction approaches and their applications, additional information can be found in the literature [9, 10, 68]. These hybrid methodologies represent promising avenues for advancing RUL prediction capabilities, offering more accurate and robust prognostic solutions for a wide range of engineering systems and applications.

2.8.2 Segmentation of long-term health index data

The calculation of the RUL should begin with the initiation of the unhealthy stage, known as the first predicting time (FPT). This is because the observations made during the healthy stage do not provide significant information regarding the degrading trend of the unhealthy stage. Hence, rather than exerting excessive effort to utilize intricate models, such as those with switching regimes or non-linear trends, a more feasible and evaluative approach may involve dividing the health index into distinct health stages (HS) and predicting the RUL based on the final stage. Typically, while HS is evaluated, the health index is compared to a limit value, which is usually given by the manufacturer. This limit value represents the threshold at which the HS transitions from a "Good Condition" (healthy stage) to the "Warning" (degradation stage) and from a "Warning" to the "Alarm" stage (critical stage). Regrettably, in numerous cases, the limit values or anticipated useful life are unknown, a common occurrence in the raw materials sector [5], particularly when the machine is unique. Furthermore, it should be noted that most of these thresholds are established by manufacturing sectors and are determined according to specific requirements for working and environmental circumstances, which may become invalid when situations alter.

Recently, numerous scholarly articles have been published on the identification of change points in time series data across various domains, including finance, medicine, and meteorology, please see the following references [69, 70, 71, 72, 73, 74].

Additionally, it is worth mentioning that this procedure is referred to as regime-switching point identification or signal segmentation in alternative circles. Signal segmentation is a commonly used technique in signal processing applications to divide the original data into segments that are similar or to extract the underlying pattern. Gasior et al. [75] employed segmentation techniques to isolate shocks from vibrations in sieving screens. Kucharczyk et al. [76] employed stochastic modeling to segment seismic signals. Grzesiek et al. [77] introduced an approach for identifying regime changes, namely when regime A undergoes a seamless transformation into regime B. Furthermore, several studies have been published within the PHM community, focusing on the categorization of health index into various HS. Several of these publications categorize HI into two distinct stages. Alkan et al. [78] introduced an approach for diagnosing problems in electromechanical systems. The method relies on a variance-sensitive adaptive alarm threshold and principal component analysis (PCA). Fink et al. [79] described the prediction of RUL as a two-stage classification process aimed at identifying the machine's condition after a specified time interval. Hu et al. [80] employed a one-class support vector machine and a Gaussian threshold model to perform condition monitoring on turbo-pumps.

The two-stage division is only valuable for cases where the degradation trends of machinery in the unhealthy stage are consistent and can be represented utilizing a single degradation model. However, the degradation trends of machinery may change as a result of different fault patterns or operational situations. In light of this situation, the detrimental phase should be divided into distinct stages according to varied patterns of degeneration.

A few of these research divided HI and spectra into several stages, with changing points detection approaches. Kimotho et al. [81] divided the degradation trend into five stages based on the changes in frequency amplitude in the power spectra density. Sutrisno et al. [82] categorized the degradation of the bearing into several phases using anomaly detection of frequency spectra. Hu et al. [83] divided the concept of HI into four stages by identifying the points at which confidence levels change.

Machine learning methods are commonly employed for long-term data analysis in tasks such as categorization, defect detection, and prognosis. These techniques have been applied in various studies, including those by [5, 84, 85, 86, 87]. In this scenario, unsupervised classification, namely clustering, is necessary due to the varying settings under which each machine operates, making data labeling a challenging endeavor. Therefore, the clustering method has significant potential to partition long-term data into regimes. The researchers in [85] introduced a long-short-term memory (LSTM) network combined with a clustering technique to forecast the RUL in many stages. Singh et al. [86] proposed an adaptive data-driven model-based method for identifying regime-changing locations using K-means clustering.

Discrete state transition models, such as hidden Markov models (HMMs) [88, 89, 90], and dynamic state-space models [91, 92, 93], are commonly used to divide degradation processes into many stages. However, the majority of the research cited

was conducted with the premise that observation noise follows a Gaussian distribution. However, in many real-world applications, particularly in the field of PHM, this assumption is not appropriate. Impulsive noise is frequently detected and can be accurately characterized by a heavy-tailed distribution.

Due to the random nature of the degradation process seen in the HI data, conventional algorithms described in the literature may not be able to perform the segmentation procedure accurately [94, 95, 96].

Chapter 3

Experimental data

3.1 Experimental data

In this section, the experimental dataset utilized in this thesis will be thoroughly examined

3.1.1 FEMTO dataset

Introduction of the dataset

The dataset was released to the public during the IEEE International Conference on PHM in 2012 and was provided by the Franche-Comté Electronics Mechanics Thermal Science and Optics–Sciences and Technologies Institute, as stated by [97]. Consisting of 17 run-to-failure datasets of rolling element bearings obtained from the PRONOSTIA platform, as shown in Fig. 3.1, this dataset was used to perform accelerated degradation tests, effectively simulating the equivalent of several years of bearing wear in a few hours. This was accomplished by providing a significant magnitude radial force that exceeded the maximum dynamic load capacity of the bearings. During testing, the rotational speed of the bearings remained consistent. The data collection process was made easier by using two accelerometers and a thermocouple, which recorded vibration signals and bearing temperatures. Bearing failure was determined to have occurred when the magnitude of the vibration signal was greater than 20 g.

This dataset specifically captures instances of bearing degradation that occur naturally, without any preexisting flaws or artificial manipulations. Adding to the complexity of the dataset is the restricted accessibility of only two training units for each operational situation. In addition, various failure patterns and lifespans distinguish distinct units, even when exposed to identical conditions. The variation between training and testing units exacerbates the issue of predicting the RUL.

The vibration signals in the dataset have poor frequency resolution because each sample covers a time interval of 0.1 s, resulting in a frequency resolution of 10 Hz. Therefore, traditional fault diagnostic techniques that depend on intricate frequency analysis are not suitable for this particular dataset.

In addition, the dataset demonstrates the spread of initial faults from one component to other components through frequent contact, leading to the simultaneous emergence of several fault patterns, as seen in Fig. 3.2. This behavior highlights the appropriateness of the dataset for dealing with complex predictive tasks involving several faults.

This dataset is commonly used for health index construction [98, 99, 100, 101, 102, 103, 15, 104], healthy stage evaluation [105, 106, 107, 108, 5], and RUL prediction [109, 110, 111, 31, 112, 32].

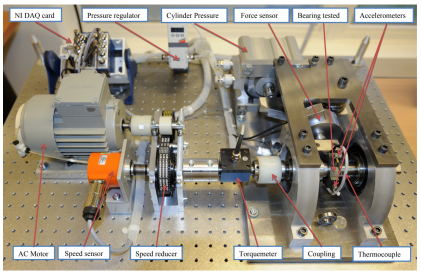


FIGURE 3.1: FEMTO test rigs [97].

FIGURE 3.2: Different faults of rolling element bearings [97].

In this thesis, the proposed methodology is applied to the bearing labeled $1-1$ within the FEMTO dataset. The data in this dataset have been gathered under specific operating conditions, precisely at 1800 rpm and with an applied load of 4000 N. For each set of vibration data, the RMS is calculated, which is subsequently used as the health index for our analysis, see Fig. 3.3.

Our selection of this particular case study stems from the fact that this case is a good example of a three-stage degradation model, as shown in Fig. 3.3. This model exhibits distinct stages: initially, a consistent amplitude of the vibration signal characterizes a healthy stage; subsequently, a gradual linear increase in amplitude signifies the onset of degradation; finally, a dramatic surge in the amplitude of the signal marks the critical stage. This progression shows a monotonic behavior, which is significant for our analysis. Furthermore, as noted in [113], the random component of this case study adheres closely to a Gaussian distribution. This observation supports our objective of demonstrating the effectiveness of our proposed model in a real world scenario where the distribution of random parts is approximately Gaussian.

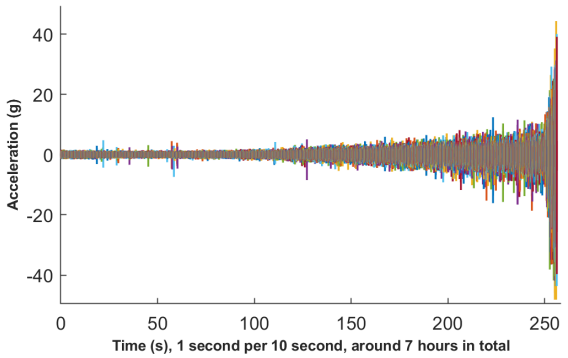


FIGURE 3.3: Raw vibration data for case study bearing $1-1$, each set of color is related a set of raw vibration data.

3.1.2 IMS dataset

The data collection in issue was sourced from the intelligent maintenance systems (IMS) center at the university of Cincinnati [114], and it is accessible through NASA's prognostic data repository [115]. This dataset consists of three separate subsets of bearing deterioration tests. Each test consisted of installing four Rexnord ZA-2115 double row bearings on a shared shaft, as shown in Fig. 3.4. Vibration signals were captured by placing accelerometers on bearing housings, and a specialized oil circulation system was designed to guarantee proper lubrication of the bearings. To facilitate debris collection, a magnetic stopper was incorporated into the oil feedback pipe. The test execution was meant to be adaptive, using an electrical switch that would automatically stop the process when the amount of trash reached a specific limit. After the tests were completed, the bearings were thoroughly examined and the identified fault patterns carefully documented.

Vibration signals possess a high level of frequency resolution, making them very suitable for diagnosing faults using frequency analysis techniques. As stated in the data instructions [115], each file contains 20480 data samples that were taken at a rate of 20 kHz. As a result, operators can obtain important characteristics in the frequency domain to track the deterioration of particular bearing components, such as rollers, outer ring, and inner ring. After the tests were completed, the researchers saw and documented precise fault patterns for each bearing. These fault patterns are shown in Fig. 3.5, providing valuable information for researchers to investigate the connections between distinct fault patterns and their associated deterioration trends.

This dataset has been extensively used in various publications concerning segmentation [116, 117], RUL prediction [118, 119, 120, 121, 12] and condition monitoring [114].

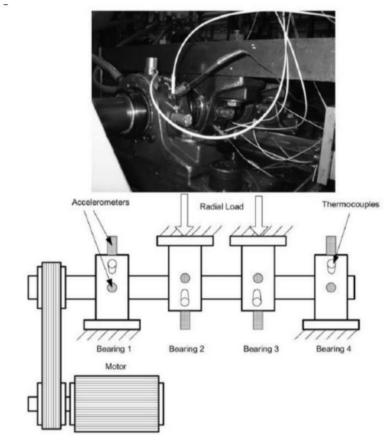


FIGURE 3.4: IMS test rig [114].

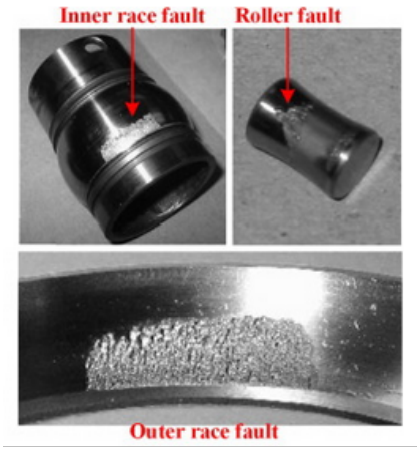


FIGURE 3.5: Different faults of rolling element bearings [114].

The IMS data collection consists of three separate subsets. In this thesis, subset 3 is selected and we chose to examine the vibration of bearing number 3 as a case study, as shown in Fig. 3.6. Subset 3 collects data during a recording period that starts on April 4, 2004, at 09:27:46, and ends on April 4, 2004, at 19:01:57. The dataset consists of vibration data taken from four channels. The channel arrangement is as follows: bearing 1 corresponds to channel 1, bearing 2 corresponds to channel 2, bearing 3 corresponds to channel 3, and bearing 4 corresponds to channel 4. The

information is documented at a 10-minute frequency in ASCII file format. The experiment ends with a failure event in the outer race of bearing number 3, as seen in Fig. 3.5.

This specific case from the IMS dataset is selected because it has unique characteristics, particularly the unusually short period between deterioration and critical stages. This rapid transition could potentially be classified as a two-stage model rather than a multistage model. This particular case also presents an extra situation in which the noise distribution closely resembles a Gaussian distribution, hence increasing its significance for our research.

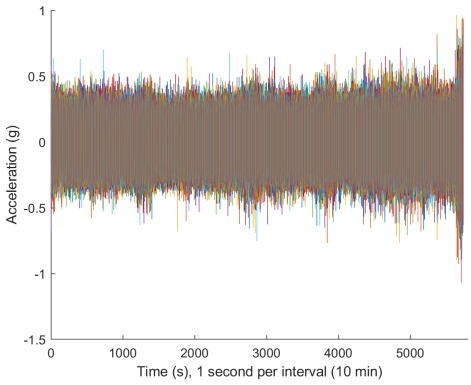


FIGURE 3.6: Raw vibration data for IMS dataset: subset 3, bearing 3, each set of color is related a set of raw vibration data.

3.1.3 Wind turbine dataset

The wind turbine dataset consists of sensor data obtained from the high-speed bearing shaft of a 2.2 MW wind turbine, as shown in Fig. 3.7. This dataset includes measurements of the energy of the bearing's inner race, taken at 10-minute intervals during a span of about 50 days. The specific bearing being examined exhibits an inner race defect and is classified as SKF 32222 J2 tapered roller bearings. This specific tapered roller bearing (TRB) has an outer diameter of 200 mm, an inner diameter of 110 mm, and a total length of 56 mm. It is equipped with 20 rolling components arranged at a 16 ° taper angle and has an approximate weight of 20 pounds. The bearing in issue is supported by two pillow blocks and is supplied with a load cell to quantify the force exerted on the bearing. The wind turbine functions at varied frequencies ranging from 2 to 100 Hz, with a load cell that can measure loads up to 1000 pounds. However, most of the tests were carried out at loads of either 150 or 300 pounds. To reduce the probability of severe gearbox failure, the maximum test load applied was kept at 50% of the power rating. More detailed information on this dataset can be found in the reference cited as [122].

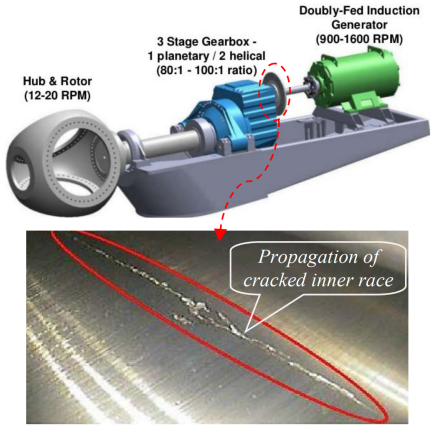


FIGURE 3.7: Wind turbine test rigs [122].

The inner race energy is selected as the representative health index for the wind turbine dataset in our suggested approach. To obtain more comprehensive information on the process of extracting health indicators, please refer to the publication by Bechhoefer et al. [123] . Specifically, as shown in Fig. 3.8, this dataset is distinguished by a conspicuous presence of outliers. Therefore, it is suitable to classify this dataset as having a trend affected by non-Gaussian noise.

Moreover, the health index experiences fluctuation that can be ascribed to dynamic factors such as load variations or even phenomena such as self-healing. The intricate behavior, which is observed in real-world situations, poses a significant difficulty for the tasks of segmentation and predicting RUL. The intricate dynamics of the system highlight the inherent challenge in accurately characterizing it for the actual implementation of predictive maintenance.

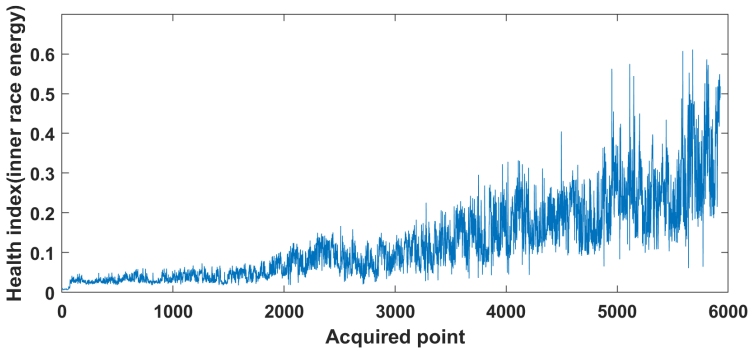


FIGURE 3.8: Wind turbine health index [122].

3.2 Synthetic degradation model

Within the realm of PHM, a widely accepted belief centers on the path that HI values follow as the degradation process unfolds. Several current approaches are based on the assumption that HI values demonstrate distinct patterns during the lifetime of the machine, as illustrated in Fig. 3.9.

In this thesis, the degradation process is shown as consisting of three separate regimes, as shown in panel (f) of Fig. 3.9. Each regime may have both trend and random components (see Fig. 3.10).

Regime 1 proposes a condition of no deterioration, perhaps with possible slight variations. Regime 2 is responsible for a phase of linear degradation increase that is accompanied by random fluctuations in noise (referred to as the slow degradation stage). Regime 3 signifies a rapid increase in degradation, characterized by exponential growth. This model is widely used in the literature [124, 125].

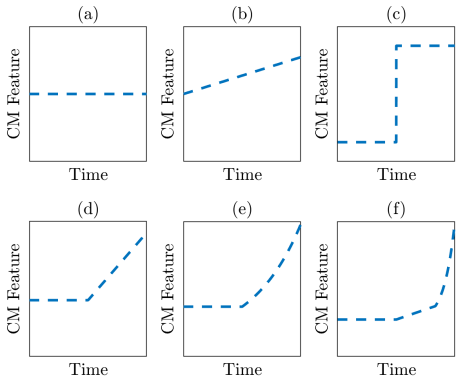


FIGURE 3.9: Types of feature variations and degradation models: (a) good condition (constant trend), (b) good to gradual wear (linear trend) (c) switch from good to bad condition, (d) good to accelerated wear (linear trend), (e) good to accelerated wear (exponential trend), (f) three regimes model (good, linear progress and exponential progress of degradation).

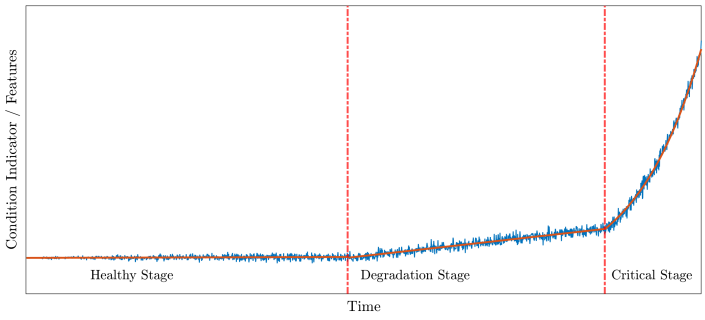


FIGURE 3.10: The preliminary model used for our analysis. The red solid line represents the theoretical trend line, two vertical dashed lines point out the change of regime (Healthy Stage/ Degradation/ Critical stage). Note constant variance of random component.

Upon initial examination of the actual HI data, it becomes evident that they display heterogeneous properties that change over time. Furthermore, these data may contain a substantial level of interdependence among random components and exhibit a distribution with heavy tails that deviate from the Gaussian distribution. In addition, our observations indicate the presence of three separate regimes, each corresponding to various stages of the process of degeneration. The data in these regimes exhibit varying properties, including differences in both deterministic components and scales. Additionally, there is a transition in the distribution of random components from Gaussian (or nearly Gaussian) to severely non-Gaussian distributions. The proposed HI is illustrated in Fig. 3.11.

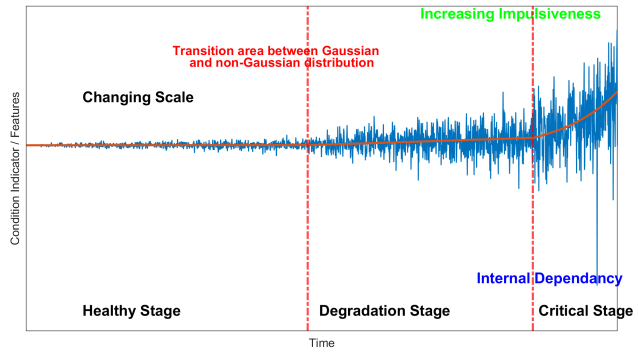


FIGURE 3.11: Long-term data variation simulated from the adopted model.

Based on the data features mentioned before, the following model is suggested for synthetic degradation data $\{S(t)\}$ for the simulation research

$$S(t) = R(t) + T(t). \tag{3.1}$$

The terms $\{R(t)\}$ and $T(t)$ refer to the random and deterministic (that is, trend) components, respectively. Both of these components are comprised of three distinct regimes, labeled regime 1, regime 2, and regime 3. These regimes correspond to three underlying states (healthy, warning, and alarm) and dictate the behavior of the process in terms of both the trend and noise scale. Suppose that we have a sample signal HI denoted as $S(1), \cdots, S(N)$. The transition point from regime 1 to regime 2 is represented as τ_1, while the transition point from regime 2 to regime 3 is represented as τ_2, where $1 < \tau_1 < \tau_2 < N$. To clarify, the signal can be divided into different regimes. Specifically, the sequence $S(1), \cdots, S(\tau_1)$ represents regime 1, the sequence $S(\tau_1 + 1), \cdots, S(\tau_2)$ represents regime 2, and the sequence $S(\tau_2 + 1), \cdots, S(N)$ represents regime 3. It can be used in the same terminology for the sequences $\{R(t)\}$ and the function $T(t)$.

The random component associated with $\{R(t)\}$ is created using the following method. To begin, the sequence of AR time series with an order of p that corresponds to the time series $\{R2(t)\}$ satisfying (3.2)

$$R2(t) - \phi_1 R2(t-1) - \cdots - \phi_p R2(t-p) = R3(t), \tag{3.2}$$

where $\{R2(t)\}$ is the stationary AR model with order $p > 0$ (AR(p)) and the polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0. \tag{3.3}$$

.

The sequence $\{R3(t)\}$ is independent and identically distributed (iid) random variables, where the scale parameter is equal to one. In the subsequent part, the simulation analysis is conducted under the assumption of two different distributions for the random component: Gaussian and α-stable, see Appendix A. Regarding the Gaussian distribution, it is assumed that $R3(t)$ follows a normal distribution with mean 0 and standard deviation 1. On the other hand, in the α-stable scenario, it is assumed that $R3(t)$ follows a α-stable distribution with parameters α, 0, $\frac{1}{\sqrt{2}}$, and 0.

To simplify matters, it is assumed that the distribution is the same in each regime. However, as stated in Table. 3.1, in reality, it may vary between different regimes. The random component of the model is described below

$$R(t) = SC(t)R2(t). \tag{3.4}$$

The function $SC(t)$ represents the scale that changes over time. The behavior of the system varies depending on the regime. In regime 1, the scale increases linearly from σ_1 to σ_2, with both values being relatively close to each other. In regime 2, the scale continues to increase linearly from σ_2 to σ_3. Finally, in regime 3, the scale grows exponentially from σ_3 to σ_4. The function $SC(t)$ is defined in the following manner

$$SC(t) = \begin{cases} a_1 t + b_1 & 0 < t \leq \tau_1, \\ a_2 t + b_2 & \tau_1 < t \leq \tau_2, \\ a_3 \exp(b_3 t) & \tau_2 < t \leq N. \end{cases} \tag{3.5}$$

The constants a_1, b_1, a_2, b_2, a_3, and b_3 are determined based on conditions $SC(1) = \sigma_1$, $SC(\tau_1) = \sigma_2$, $SC(\tau_2) = \sigma_3$, and $SC(N) = \sigma_4$.

The behavior of the deterministic component $T(t)$ varies depending on the regime. It should be noted that in regime 1, the level remains constant and is represented by the symbol c_1. Next, in regimes 2 and 3, it is examined linear and exponential functions are examined, respectively, that have the same growth characteristics as the corresponding regimes of the scale function $SC(t)$. Furthermore, it is made the assumption that the function $T(t)$ does not have discontinuities at the regime change points τ_1 and τ_2. Assuming these conditions, the deterministic component of the signal can be expressed as

$$T(t) = \begin{cases} c_1 & 0 < t \leq \tau_1, \\ a_2 t + c_2 & \tau_1 < t \leq \tau_2, \\ a_3 \exp(b_3 t) + c_3 & \tau_2 < t \leq N. \end{cases} \tag{3.6}$$

The values of c_2 and c_3 are determined in a manner that ensures the continuity of the function $T(t)$. In Fig. 3.12, it is shown the deterministic component $T(t)$ and the scale function $SC(t)$ for the given parameter values: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$, and $c_1 = 10$. In addition, Fig. 3.13 shows the simulated signals for the sequence $\{S(t)\}$. The random component $\{R2(t)\}$ follows an AR model with parameters $p = 1$ and $\phi = 0.5$. The sequence $\{R3(t)\}$ is generated from a α-stable distribution. In this analysis, three more values of the α- parameter are examined, specifically $\alpha \in \{1.95, 1.9, 1.85\}$. It is evident that in the case of a Gaussian-distributed signal, it cannot detect significant outliers. However, in the case of a non-Gaussian heavy-tailed distribution, strong spikes can be observed in the data. As the value of α decreases, the likelihood of experiencing higher impulses in the signal increases.

Finally, the important features of suggested model are described in Table 3.1.

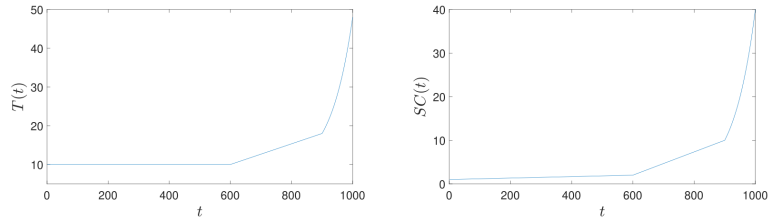


FIGURE 3.12: The deterministic component $T(t)$ and the scale function $SC(t)$ for the following values of the parameters of the model: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$ and $c_1 = 10$. The left panel is shown the determinstic part of simulated HI and the right panel is demonstrated the scale of the random part.

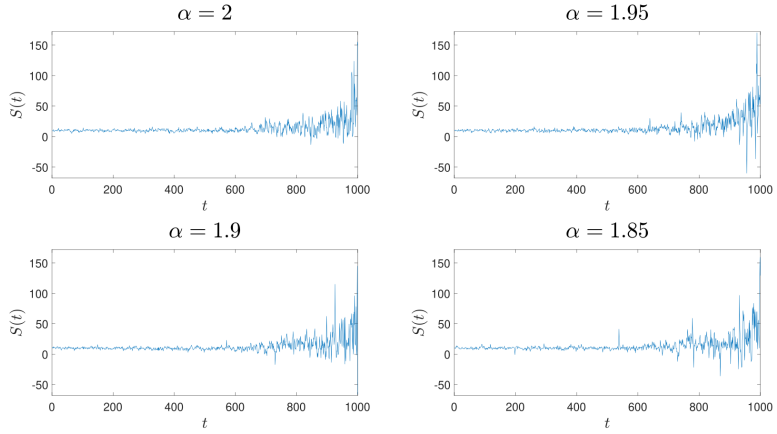


FIGURE 3.13: The exemplary data from the model with deterministic component $T(t)$ and scale function $SC(t)$ presented in Fig. 3.12. The random components of the model correspond to the AR(1) model with $\phi = 0.5$. Four cases of distributions are considered: Gaussian (α-stable with $\alpha = 2$), and non-Gaussian α-stable with $\alpha = 1.95, 1.9$ and $\alpha = 1.85$.

TABLE 3.1: Main characteristics of the data for three regimes indicated in Fig. 3.11.

	regime 1	regime 2	regime 3
Trend	constant	linear	exponential
Scale	nearly constant	linearly growing	exp. growing
Autodependence of random component	relatively small	significant	significant
Coefficients of the stochastic model	negligible	significant	significant
Distribution of the random component	nearly Gaussian	non-Gaussian	strongly non-Gaussian

Chapter 4

Long-term health index data modeling and identification

4.1 Introduction of long-term health index data modeling

Analyzing historical data is a crucial step in the data-driven approach to build models that accurately predict sample of HI. In Chapter 3, three open-source benchmark datasets are introduced, along with an artificial three-stage degradation model developed to generate synthetic degradation data. Continuing from the previous chapter, this section aims to introduce a framework and tools to identify the characteristics of series of health indexes. However, conventional approaches often face challenges when handling extended non-stationary data with non-Gaussian characteristics. In addition, due to the nature of the degradation process, the behavior and characteristics of the HI can completely change from one stage to another. Therefore, it is essential to segment and analyze the data on the basis of stages. This chapter presents a detailed methodology and appropriate mathematical tool specifically to handle complex health index datasets. The segmentation of long-term degradation data is discussed in detail in Chapter 5.
This framework addresses the challenge of distinguishing between deterministic and random elements, capturing the heavy-tailed and time-varying characteristics of the data, and detecting any hidden dependencies within the random part. In addition, it identifies the distribution that governs the random component.
In summary, this chapter establishes a systematic framework for analyzing data with the aim of creating models for intricate and enduring datasets. Although the proposed approach draws inspiration from predictive maintenance, it appears to be applicable to a wider range of individuals, particularly academics who deal with non-Gaussian, non-homogeneous, and time-varying data.

4.2 Methodology: a framework for long-term health index data modelling

The schematic diagram of the framework is shown in Fig. 4.1. To begin, the outline the overarching approach for the proposed framework is investigated. Next, the illustration of the techniques employed in the scheme for data that follows a Gaussian distribution (Section 4.2.1) and data that exhibits non-Gaussian behavior (Section 4.2.2) is described.

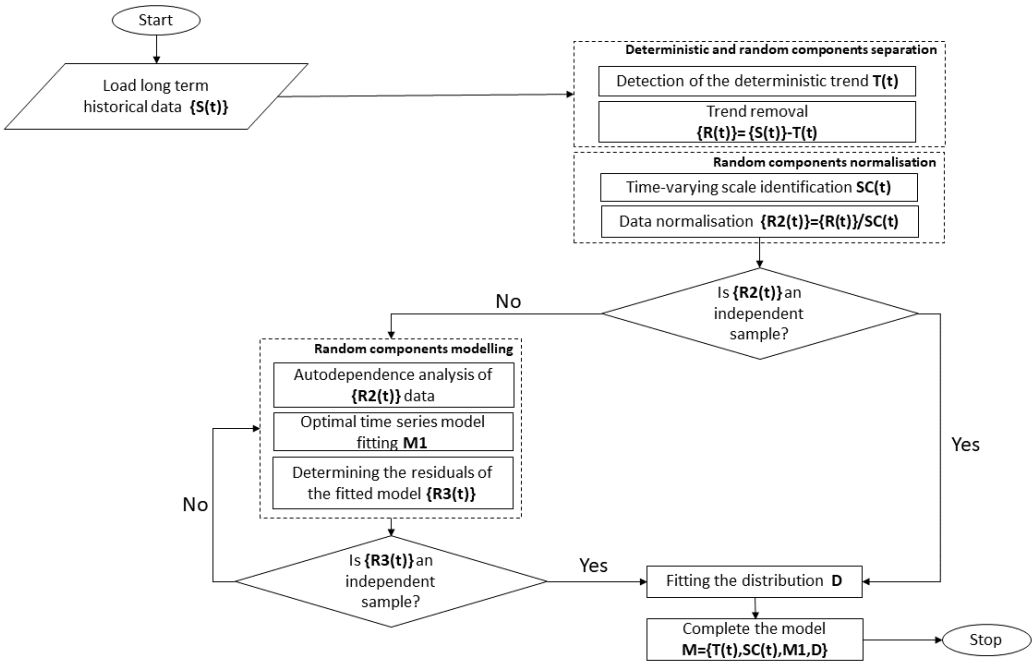


FIGURE 4.1: The framework for modelling of long-term data. Each block of this diagram is described in details in Section 4.2.

Deterministic and random components separation
Initially, we detect the deterministic component (represented as $T(t)$) in the long-term data (represented as $\{S(t)\}$). Based on our initial investigation, it appears that the data analyzed in this chapter cannot be accurately characterized by a single deterministic function. Specifically, significantly intricate scenario in which the deterministic component varies based on the condition associated with the healthy state, warning, and alarm is witnessed. Therefore, in this following, the deterministic component by calculating the empirical location measure for overlapping segments from windows of a specified length w, without segmenting the data is determined. For the measurement of location, two statistical methods are suggested. Please refer to Section 4.2.1 for data that follows a Gaussian distribution, and Section 4.2.2 for data that does not follow a Gaussian distribution. This stage is essential for seprating the random component that is sprated from the raw data. The series, written as $\{R(t)\}$, is obtained by eliminating the deterministic component.

Random component normalisation
The time series $\{R(t)\}$ is considered to represent the random component of the raw data. Our initial examination unequivocally demonstrates that the sequence $\{R(t)\}$ has a non-homogeneous structure primarily influenced by the changing scale throughout time. In the next stage of the proposed framework, the goal is to standardize the data by determining the time-dependent empirical scale parameter. This parameter is then computed for segments of length w with overlapping o for the series corresponding to $\{R(t)\}$. Consequently, the component $SC(t)$ is acquired. Afterwards, data normalization to produce the series $\{R2(t)\}$ is done, which exhibits a homogenous structure with time-constant properties. The issue of identifying the time-varying scale is addressed separately in Section 4.2.1 for the Gaussian case and in Section 4.2.2 for the non-Gaussian case.

Random component modelling

Next, the autodependence of the time series $\{R2(t)\}$ is analyzed. It is essential to undertake this stage in order to identify the model that accurately describes the signal. Once the sample autodependence measurements confirm that the signal consists of independent observations, the next step is to identify the appropriate distribution. Alternatively, it is necessary to initially select the appropriate time series model and thereafter compute its residuals, represented as $\{R3(t)\}$. Autodependence measurements might also be helpful at this stage to verify if the residuals of the fitted model can be regarded as independent observations. Ultimately, the appropriate distribution D is fitted to the sequence $\{R3(t)\}$ (or the sequence $\{R2(t)\}$). The sample autodependence measures utilized in the aforementioned stages are explained in Section 4.2.1 for the Gaussian case and in Section 4.2.2 assuming a non-Gaussian distribution. The series $\{R2(t)\}$, which is generated by normalizing the time-dependent scale parameter, demonstrates behavior that is suitable for the linear AR model. Within the framework, it refers to it as $M1$. Also, the time series $\{R3(t)\}$ is iid. In the classic approach, it is assumed that the residuals $\{R3(t)\}$ follow a Gaussian distribution. However, there are also circumstances when non-Gaussian heavy-tailed distributed noises are examined. The data examined in this chapter also demonstrate such characteristic.

Subsequently, it is suggested using a straightforward visual test to determine if a distribution fits properly and to verify if it exhibits non-Gaussian behavior. The test involves comparing the empirical tail of the distribution with the theoretical tail of the tested distribution, using the estimated parameters obtained from the random sample under consideration. Through this visual examination, it can also determine which of the tested distributions is more suitable for the analyzed data. The tail of the distribution of a random variable X is defined as $1 - F_X(x)$, where $F_X(\cdot)$ represents the cumulative distribution function (CDF) of X. The empirical tail of the random sample $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ is the complement of the empirical CDF $\hat{F}_{\mathbf{x}}(x)$, where $\hat{F}_{\mathbf{x}}(\cdot)$ is the empirical CDF for x as defined in [126]

$$\hat{F}_{\mathbf{x}}(x) = \frac{1}{n} \sum_{j=1}^{n} 1\{x_j \leq x\}, \tag{4.1}$$

where $1\{A\}$ represents the indicator function of a set A. The theoretical distributions that include the Gaussian distribution, the α-stable distribution, and the t location scale distribution are tested . The α-stable and t location scale distributions are categorized as non-Gaussian heavy-tailed distributions, see Appendix A.

In the simulation research described in Section 4.3, the Kolmogorov-Smirnov (KS) statistic is employed to verify that the tested distribution is the most appropriate one, or more suitable than the alternative distribution. The KS statistic is often used to measure the distance between the empirical distribution function of a sample and the CDF of a reference distribution, or between the empirical distribution functions of two samples. Additionally, it serves as a test statistic for the KS goodness-of-fit test

$$KS = \sup_{x} \left| \hat{F}_{\mathbf{x}}(x) - F_X(x) \right|. \tag{4.2}$$

The function $F_X(\cdot)$ represents the CDF of the theoretical distribution being evaluated. The parameters used in this CDF are calculated from the random sample $\mathbf{x}$. A low value of the KS statistic suggests that the observed distribution closely resembles the distribution being tested. For the estimation of the parameters, for all considered distributions, we use the MLE method.

In the last stage of the proposed framework, the deterministic trend $T(t)$, the time-varying scale $SC(t)$, the time series model $M1$, and the distribution D that is fitted to the independent observations of either $\{R3(t)\}$ or $\{R2(t)\}$ are combined.

In the subsequent sections, a comprehensive explanation of the algorithms employed in the framework that has been provided. As previously stated, it is necessary to replace the conventional procedures designed for Gaussian distributed models with robust ones that are better suited for non-Gaussian heavy-tailed data in all steps of the procedure.

4.2.1 Algorithms for Gaussian distributed data

Deterministic and random components separation
Within the overall structure, the initial pivotal stage involves determining the deterministic trend $T(t)$. The classical statistic commonly used as a measure of location is simply the average of a sample. In this scenario, it simply compute the moving average (MA) for overlapping segments (with an overlap of o) of a specific length w and use it as the deterministic component of the signal $\{S(t)\}$.

Random component normalisation
The conventional estimator for the scale parameter is the sample standard deviation, which is defined as the square root of the variance of the signal $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$

$$SC_{\mathbf{x}} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \overline{\mathbf{x}}\right)^2}, \tag{4.3}$$

where $\overline{\mathbf{x}}$ represents sample mean for the signal $\mathbf{x}$. Due to the presence of time-varying scale in the data, the empirical standard deviation in the Gaussian scenario for segments of length w with overlapping o is utilized. Under those circumstances, the segments are regarded as quasi-homogeneous sub-signals.

Random component modelling
Within this section , our initial step involves analyzing the autodependence of the normalized signal associated with $\{R2(t)\}$. This component is essential for verifying whether the signal can be regarded as an independent sequence or for identifying the appropriate time series model that accurately describes the data (referred to as $M1$). The autocorrelation function (ACF) is the primary measure of autodependence utilized in time series analysis. The sample version of the stationary signal $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ is defined as follows [127]

$$\hat{\rho}_{\mathbf{x}}(h) = \frac{\hat{\gamma}_{\mathbf{x}}(h)}{\hat{\gamma}_{\mathbf{x}}(0)}, \quad -n < h < n, \tag{4.4}$$

where $\hat{\gamma}_{\mathbf{x}}(\cdot)$ is the sample autocovariance function (ACVF) given by

$$\hat{\gamma}_{\mathbf{x}}(h) = \frac{1}{n}\sum_{j=1}^{n-|h|}(x_j - \overline{\mathbf{x}})(x_{j+|h|} - \overline{\mathbf{x}}). \tag{4.5}$$

Furthermore, h is referred to as a lag. When the analyzed signal follows a distribution with limited variance, such as a Gaussian distribution, the estimator $\hat{\rho}_{\mathbf{x}}(\cdot)$ is regarded as the most efficient estimate of the ACF. By visual inspection its behavior,

one can determine whether the signal in question displays long- or short-range dependence, or if it can be regarded as an independent sample. The high significance level of the $\hat{\rho}_{\mathbf{x}}(\cdot)$ function for $h > 0$ suggests that the considered signal exhibits self-dependence. The suggested framework applies the sample ACF to the Gaussian case for the signal corresponding to $\{R2(t)\}$ and $\{R3(t)\}$, which are residual series of the fitted time series model.

After confirming the dependence of the data, it can proceed to fitting the time-series model in the random component's modeling phase. In this analysis, the model of order p as defined in Eq. (3.2) is examined. In order to determine the most suitable order p, utilizing a criterion that is universally applicable and does not require any assumptions about the distribution of the data is suggested. For any value of p ranging from 1 to p_{max}, the parameters of the AR(p) model using the Yule-Walker method, specifically in the case of a Gaussian distribution are calculated. Next, the estimated model's residuals are calculated and their sample ACF is analyzed. To be more precise, the statistic is calculated and defined as

$$K_{\mathbf{x}}(p) = \max_{h=1,\cdots,h_{max}} |\hat{\rho}_{\mathbf{x}}(h)|^2. \tag{4.6}$$

The function $\hat{\rho}_{\mathbf{x}}(\cdot)$ is defined in Eq. (4.4) for the residual series obtained from the fitted AR(p) model. The value h_{max} represents the set maximum lag value. The optimal order, denoted as p_{opt}, is chosen when the statistic $K_{\mathbf{x}}(p_{opt})$ reaches its smallest value. This straightforward approach is easily understood. The ideal model is one in which the residuals can be regarded as an independent sample. This means that the sample ACF for residuals at lag h>0 is either zero or extremely close to zero.

After identification of the proper order$p$,the parameters of the AR(p) model are estimated. The Yule-Walker method is the standard strategy for finite-variance models, and it relies on the autocovariance function (ACVF) of the time series model. The algorithm replaces the theoretical ACVF with the empirical ACVF stated in Eq.(4.5). Additional information regarding the Yule-Walker algorithm for the AR model with a finite-variance distribution can be found in the [127].

4.2.2 Algorithms for non-Gaussian heavy-tailed distributed data (robust approach)

Deterministic and random components separation

If the data shows non-Gaussian behavior, the MA method may not be enough for accurately identifying the deterministic component since it is affected by extreme observations. This is particularly crucial when the window length w is quite short and there are isolated outliers in the signal. The presence of big values can have a considerable impact on the sample mean, perhaps causing the observed trend to deviate from the expected theoretical trend. Therefore, when dealing with non-Gaussian data, it is advisable to utilize a more robust statistical measure such as the empirical location measure. This chapter utilizes the moving median (MM) technique to analyze segments extracted from windows of a specified length, denoted as w, with a specified amount of overlap, denoted as o.

Random component normalisation

If the data display non-Gaussian behavior, the theoretical variance (and therefore standard deviation) can be infinite, such as in the case of the α-stable distribution with $\alpha < 2$. Therefore, in this scenario, utilizing the robust estimator of the scale parameter for the signal $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$, which is determined by the kth order

statistic [128, 129, 130] is suggested

$$SC_{\mathbf{x}}^{Q} = d\{|x_i - x_j|; i < j\}_{(k)}, \tag{4.7}$$

where d is a constant value and $k = \binom{h}{2} \approx \binom{n}{2}/4$, where $h = [n/2] + 1$ and $[\cdot]$ denotes the integer part.

Random component modelling
The conventional ACF is not suitable for identifying autodependence in non-Gaussian heavy-tailed data. Therefore, utilizing the robust form of the sample ACF as defined in Eq. (4.4) for the classic (Gaussian) scenario is suggested . Given the sorted signal $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$, two vectors: vector $\mathbf{u}$, which includes the first $n - h$ observations, and vector $\mathbf{v}$, which includes the last $n - h$ observations are calculated. The robust sample ACF is defined as follows [129]

$$\hat{\rho}_{\mathbf{x}}^{Q}(h) = \frac{(SC_{\mathbf{u}+\mathbf{v}}^{Q})^2 - (SC_{\mathbf{u}-\mathbf{v}}^{Q})^2}{(SC_{\mathbf{u}+\mathbf{v}}^{Q})^2 + (SC_{\mathbf{u}-\mathbf{v}}^{Q})^2}, \tag{4.8}$$

where $SC_{\mathbf{x}}^{Q}$ is defined in Eq. (4.7). The robust version of ACVF is given by

$$\hat{\gamma}_{\mathbf{x}}^{Q}(h) = \frac{1}{4}\left[(SC_{\mathbf{u}+\mathbf{v}}^{Q})^2 - (SC_{\mathbf{u}-\mathbf{v}}^{Q})^2\right]. \tag{4.9}$$

In order to determine the most suitable order p for the AR model in the non-Gaussian scenario, a new approach that can be seen as a robust adaptation of the methodology previously used in the Gaussian situation is utilized. Specifically, for a given range of values for p ($p = 1, 2, \cdots, p_{max}$), the statistic $K_{\mathbf{x}}(p)$ as specified in Eq. (4.6) is calculated. However, the sample ACF defined in Eq. (4.4), is replaced here by robust variant as given in Eq. (4.8). In addition, the specific methodology outlined below, which is a robust variation of the Yule-Walker method, to estimate the parameters of the AR model is suggested. The optimal order p_{opt} is determined by selecting the argument for which the statistic $K_{\mathbf{x}}(p)$ is minimized.

Ultimately, the values of the optimal AR(p) model's parameters using the resilient variation of the Yule-Walker approach discussed in Section 4.2.2 are calculated. However, in this case, the robust estimator of the sample ACVF, which is specified in Eq. (4.9) is utilized.

4.2.3 Summary

In Table 4.1, a brief summary of the framework discussed above is provided. Here, the key distinctions between the suggested methods for the classical (designed for data that follows a Gaussian distribution) and robust (designed for data that follows a non-Gaussian heavy-tailed distribution) approaches is outlined.

TABLE 4.1: The summary of the framework for data modelling - classical and robust approaches.

	classical approach	**robust approach**
Deterministic and random component separation		
Deterministic component identification	moving average	moving median
Random component normalisation		
Time-varying scale identification	sample standard deviation	sample robust scale parameter
Autodependence identification	sample ACF	sample robust ACF
Random component modelling		
Identification of the model's order	based on sample ACF	based on sample robust ACF
Estimation of the model's parameters	classical Yule-Walker approach	robust Yule-Walker approach
Identification of the residuals autodependence	sample ACF	sample robust ACF
Testing distribution of the residual series	Gaussian distribution testing	selected non-Gaussian distribution testing
Distribution fitting	MLE method for Gaussian distr.	MLE method for identified non-Gaussian distr.

4.3 Validation of the proposed approach by simulated data analysis

In this section, the efficiency of the proposed methodology for simulated signals from the model described in Section 3.2 is presented. The parameters used in the model are as follows: $\tau_1 = 600$, $\tau_2 = 900$, $N = 1000$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 10$, $\sigma_4 = 40$ and $c_1 = 10$. Moreover, the random part corresponds to the AR(1) model with $\phi = 0.5$. In the simulation study and real data analysis, $t = 1, 2, \cdots, N$ denotes the time point (number of observation).

The sequence $\{R3(t)\}$ is analyzed under four different distribution scenarios: one Gaussian distribution and three non-Gaussian α-stable distributions with α values of 1.95, 1.9, and 1.85. In each of the considered cases, it is simulated 100 datasets from the proposed model. The exemplary signals are presented in Fig. 3.13.
We apply the methods given in Section 4.2 to each distribution case of the $\{R3(t)\}$ sequence and to each simulated dataset. The window length for trend identification is assumed to be 101 observations, whereas for scale identification it is assumed to be 50. These values have been determined by experimentation, although their selection is not crucial. For each phase of the methodology, the classical approach specifically designed for Gaussian distributed signals are employed. Please refer to the left panels of Figs. 4.2 to 4.5 for more details. The middle panels of Figs labeled as 4.2 to 4.5 illustrate the robust technique specifically designed for signals with non-Gaussian heavy-tailed distributions. The Figs labeled as 4.2 to 4.5. The findings for each of the simulated datasets are presented in the left and middle panels using grey lines, whereas the equivalent theoretical functions are shown using a black line.

The fourth row displays the empirical tails $(1 - \hat{F}_{\mathbf{x}}(x))$ for the residual series $\{R3(t)\}$ obtained using the Yule-Walker method. The left panel shows the classical version, while the middle panel shows the robust version. The black lines represent the tails of the theoretical (tested) distribution $(1 - F_X(x))$.

In the final columns of Fig. 4.2-4.5, the statistics that allow us to parameterize and compare the outcomes of conventional and robust algorithms are shown. In the first and second rows of Fig. 4.2-4.5, the root mean square errors (RMSE) for the identified trends (first rows) and scales (second rows) independently for each $t = 1, 2, \cdots, N$ are calculated. The RMSE is defined as the square root of the mean

squared difference between the theoretical value and its estimator. RMSE is a commonly employed metric for quantifying the disparities between anticipated values (in this case, estimated suitable functions) and the theoretical values (depicted by black lines on the plots). The RMSE is a metric that combines the sizes of the errors in predictions for different data points into a single measure of how well the predictions perform. It is always greater than or equal to zero, and a value of zero (which is rarely attained in real-world scenarios) would imply a perfect match. Typically, a lower RMSE is preferable to a bigger one. The RMSE is computed for each time point $t = 1, 2, \cdots, N$ individually across the datasets. The boxplots of the estimators for the ϕ parameter, generated using the Yule-Walker methodology and based on the sample (or robust sample) ACVF, are presented in the third rows of Figs. 4.2-4.5. Assuming $\phi = 0.5$, the boxplot that aligns more closely with the theoretical value indicates a more efficient strategy.
In descriptive statistics, a boxplot is a graphical representation that displays the central tendency, spread, and skewness of numerical data using quartiles. Boxplots are a type of graphical representation that do not rely on specific assumptions about the statistical distribution of a population. They are used to show the variability in samples taken from a statistical population. The spacings within each subsection of the boxplot represent the extent of dispersion and skewness of the data, typically stated using the five-number summary. The boxplot is frequently employed as a means of graphically comparing different methodologies. The boxplot displaying a median that is closer to the theoretical value and a smaller box length signifies the method's higher efficiency.

Finally, in the fourth rows and right panels of Fig. 4.2-4.5, the KS statistic is presented, which is specified in Eq. (4.2), for the distribution being tested, as determined by the CDF $F_X(x)$. A lower value of the KS statistic suggests that the observed distribution of the sample is more similar to the theoretical distribution being tested.

In Fig. 4.2, the outcomes for the Gaussian distributed model is represented. Both the classic and robust techniques get similar outcomes in this scenario. The RMSE values for trend and scale identification are similar. Furthermore, the estimators of the ϕ parameter obtained by Yule-Walker techniques exhibit a high degree of similarity. The KS statistics for testing Gaussian distribution on the series $\{R3(t)\}$ are small and have similar values. This suggests that the use of classical techniques is effective for Gaussian distributed signals.

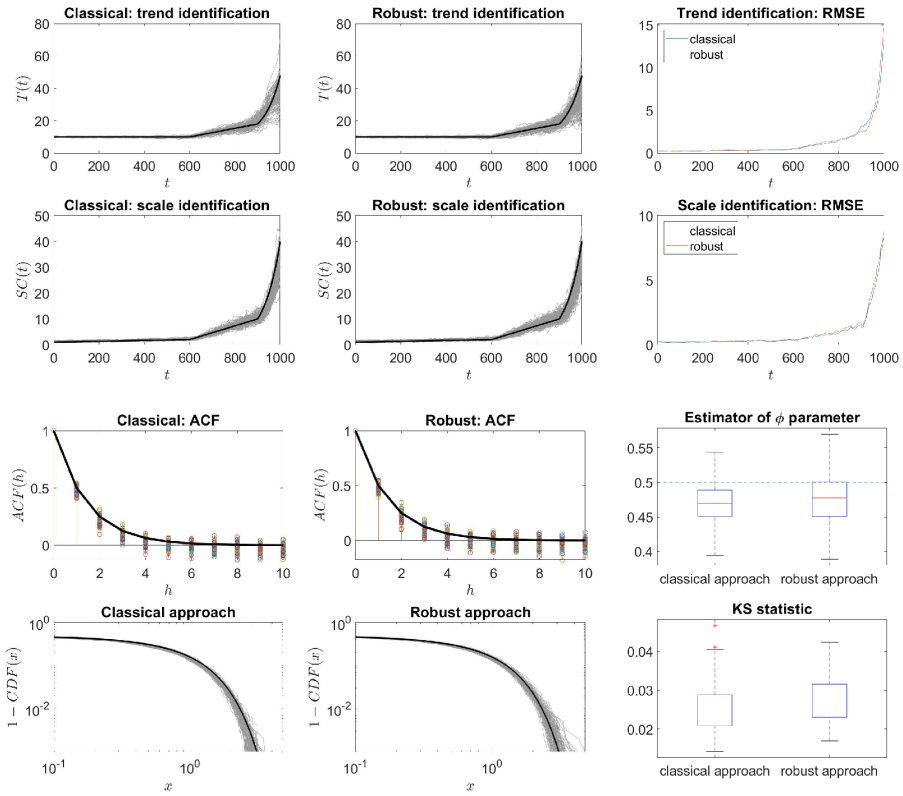


FIGURE 4.2: The outcomes of the proposed procedures for simulated data are organized across four rows, representing trend, scale, AR model, and distribution tail identification, respectively. These are presented in three columns: the left column indicates the classical approach tailored for Gaussian distributed signals, the middle column represents the robust approach suited for non-Gaussian heavy-tailed distributed signals, and the right column illustrates the evaluation of results through RMSE for trend and scale, alongside boxplots for random components. Specifically, the random component $R3(t)$ is distributed as Gaussian (i.e., α-stable with $\alpha = 2$).

The scenario shifts when the analyzed signals display non-Gaussian heavy-tailed characteristics, specifically when the random component $\{R3(t)\}$ follows an α-stable distribution with $\alpha < 2$. The appropriate situations are illustrated in Figs. 4.3-4.5.

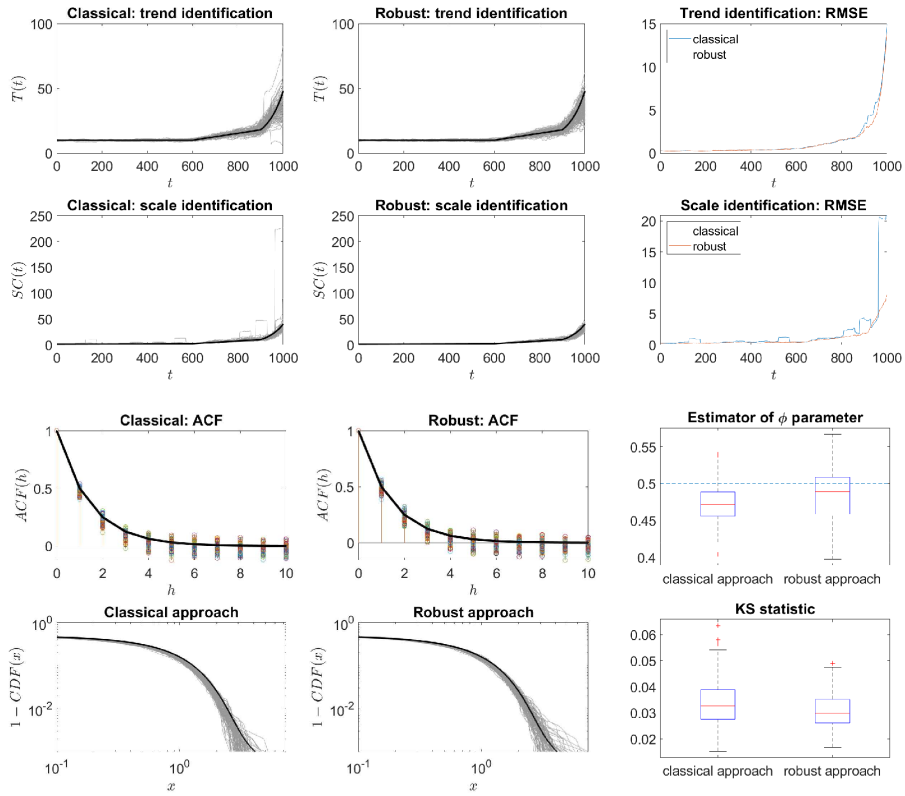


FIGURE 4.3: The outcomes of the proposed procedures for simulated data are delineated across four rows, representing trend, scale, AR model, and distribution tail identification, respectively. The comparison is presented in three columns: the left column signifies the classical approach tailored for Gaussian distributed signals, the middle column represents the robust approach suited for non-Gaussian heavy-tailed distributed signals, and the right column illustrates the evaluation of results via RMSE for trend and scale, alongside boxplots for random components. Specifically, the random component $R3(t)$ is distributed according to the α-stable distribution with $\alpha = 1.95$.

It is evident that the robust algorithms are more efficient than the traditional approaches. The methods specifically designed for non-Gaussian heavy-tailed distributed signals yield the lowest RMSE values for trend and scale identification. Furthermore, the estimators of the ϕ parameter are in closer proximity to the theoretical value $\phi = 0.5$, indicating that the robust technique performs better than the standard Yule-Walker methodology.

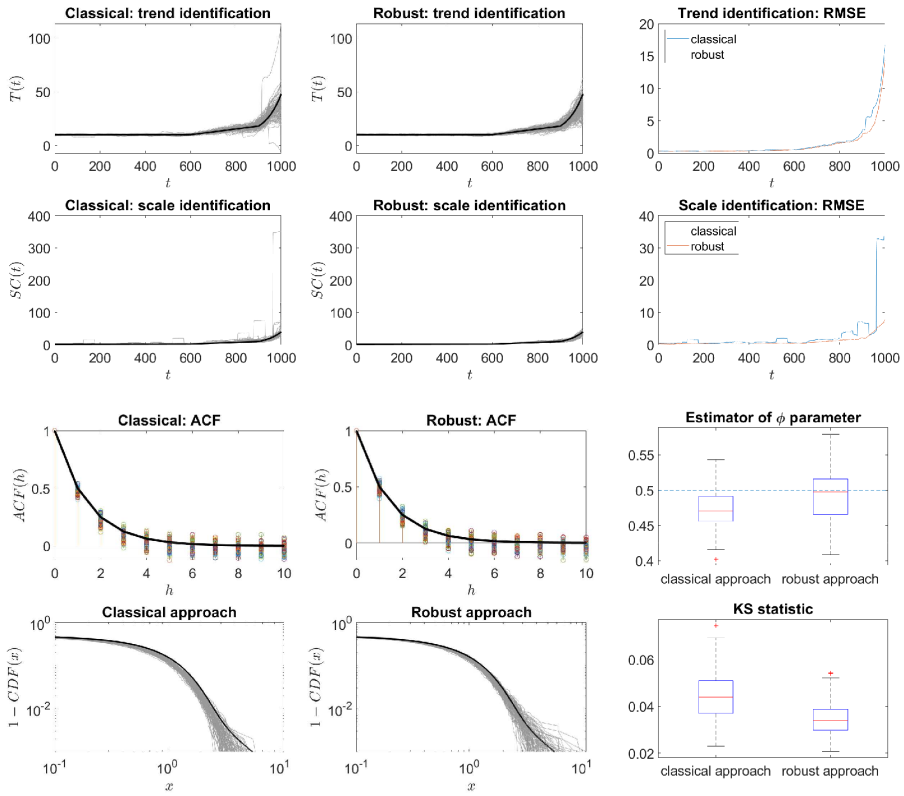


FIGURE 4.4: The outcomes of the suggested methods for simulated data are presented in rows 1-4, representing trend, scale, AR model, and distribution tail identification, respectively. The left column is dedicated to the classical technique, which is suitable for signals that follow a Gaussian distribution. The middle column is dedicated to the robust approach, which is designed for signals that have a non-Gaussian heavy-tailed distribution. The right column evaluates the results using the RMSE for both the trend and scale, as well as box-plots for the random components. The random component $\{R3(t)\}$ follows a α-stable distribution with a value of α equal to 1.9.

Finally, the comparison is made with the tested distributions, namely the α-stable distributions with α values of 1.95, 1.9, and 1.85, respectively. These distributions are shown in the fourth rows of the Figs. The Figs. labeled as 4.3 to 4.5 provide unambiguous evidence that the observed distribution of the signal associated with the $\{R3(t)\}$ sequence matches the distribution being tested when the robust version of the technique is used. The KS statistic has a smaller magnitude in the robust technique compared to the traditional methodology. The more pronounced the effectiveness of the robust algorithms becomes as the α value decreases. The simulation analysis provides clear evidence supporting the use of dedicated algorithms for heavy-tailed dispersed signals. Furthermore, the robust algorithms can be regarded as universally applicable, as they also demonstrate effectiveness with Gaussian data.

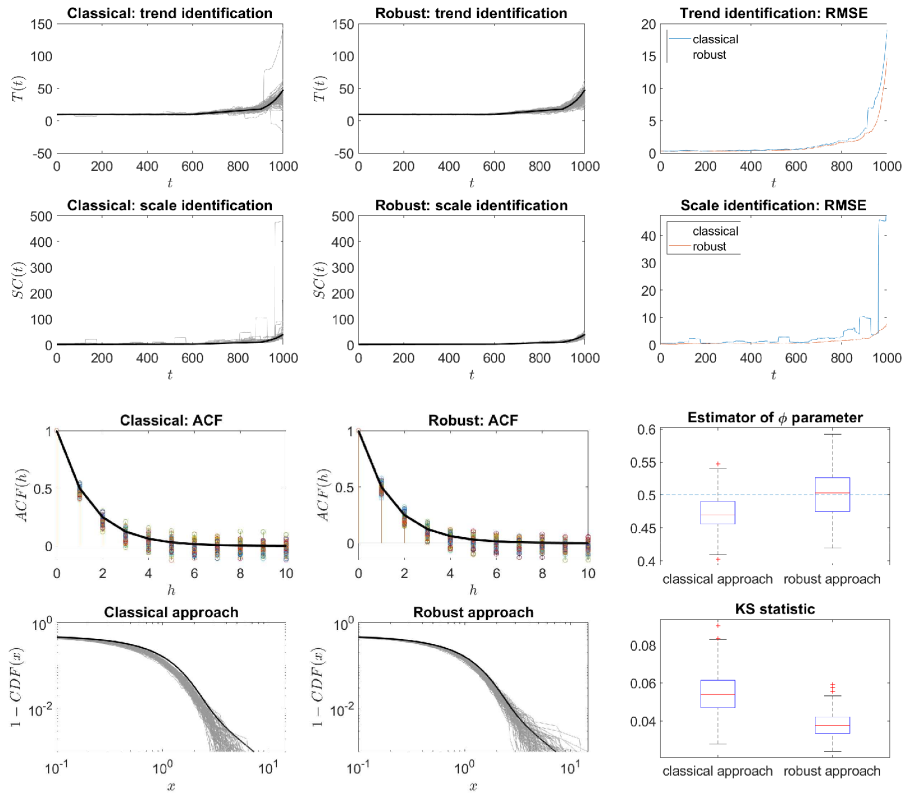


FIGURE 4.5: The simulated data was analyzed using proposed methodologies to identify trends, scales, AR models, and distribution tails. These results are presented in rows 1-4. The left column represents the traditional approach, which is designed for signals that follow a Gaussian distribution. The center column represents the robust approach, which is designed for signals that have a non-Gaussian heavy-tailed distribution. The right column is used to evaluate the results using the RMSE for both trend and scale, as well as boxplots for the random components. The random component $\{R3(t)\}$ follows a α-stable distribution with a value of α equal to 1.85.

4.4 Real data analysis

In this section, the novel methodology using established real-world datasets (FEMTO data and wind turbine data) is implemented. As mentioned, these datasets serve as standard benchmarks in academic research and competitions, each exhibiting specific noise characteristics. For both datasets, the robust approach as the proper methodology for Gaussian and non-Gaussian-distributed time series is used.

4.4.1 Results of applying the proposed framework to FEMTO dataset

The findings related to the FEMTO data are shown in Fig. 4.6. Panel (a) illustrates the identification of the trend component using a window of 51 observations, while in panel (c), the scale is identified using a window of 100 data points. In particular, the estimated trend displays varying patterns over time: initially remaining relatively constant, transitioning into a linear function, and ultimately exhibiting characteristics akin to an exponential function. This trend component aligns with the simulated model described in Section 3.2. Upon removing the deterministic trend, the random component is obtained (panel (b)), revealing non-homogeneity and varying scales within the sequence. These preliminary results validate our hypothesis, notably the observation that the scale of the random component increases over time, necessitating data normalization before further analysis.
The identified scale function, shown in panel (c), demonstrates non-linear growth in the scale of the random component, particularly evident in the final regime where noise magnitudes significantly escalate, potentially impacting prognosis quality toward the end of the lifetime curve. Panel (d) displays the random element that follows the normalization process, representing the sequence $R2(t)$. The empirical robust ACF presented in panel (e) clearly indicates short-range dependency in the data. However, the relatively modest values of the empirical robust ACF suggest insignificant coefficients in the model, attributable to nearly negligible self-dependence in the data derived from a test rig under constant load/speed conditions.
In the following, AR model will be fitted to the data, determining the optimal order of $p = 10$. The coefficients of the model are relatively small: $\phi_1 = 0.1683, \phi_2 = 0.0385, \phi_3 = 0.0114, \phi_4 = 0.004, \phi_5 = 0.0093, \phi_6 = -0.0412, \phi_7 = -0.0378, \phi_8 = 0.0034, \phi_9 = -0.0551, \phi_{10} = 0$. Panel (f) displays the residual series for the sequence $\{R3(t)\}$, with its robust empirical ACF depicted in panel (g), indicating independent observations. Comparison of actual and theoretical tails in panel (h) suggests that the residual series follows a t location-scale distribution with $\nu = 8.57$, as its empirical tail closely resembles that of the t location-scale distribution rather than the α-stable distribution. This observation rejects the Gaussian distribution assumption, further supported by panel (i), which compares the normalized histogram of the residual signal with PDF of theoretical distributions.
This outcome validates our assumption that the random component does not adhere to a Gaussian distribution, though the degree of non-Gaussian remains relatively insignificant.

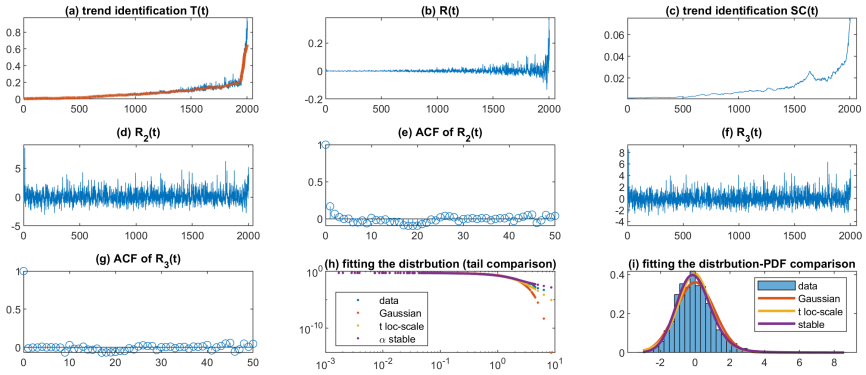


FIGURE 4.6: FEMTO data modelling - the robust approach. Results of each step of the proposed framework: (a) identified trend $T(t)$, (b) random component $\{R(t)\}$, (c) identified scale $SC(t)$, (d) normalized random component $\{R_2(t)\}$, (e) autocorrelation of normalized random component, (f) residuals of AR model, (g) autocorrelation of the residual series, (h) comparison of fitted tails, (i) comparison of fitted distributions PDF.

In Table 4.2 the characteristics identified for the FEMTO dataset is summarized.

TABLE 4.2: Identified characteristics for the FEMTO data set.

Trend	time-varying
Scale	time-varying
Autodependence of random component	negligible
Coefficients of AR model	negligible
Distribution of the random component	t location-scale with $\nu = 8.57$.

4.4.2 Results of applying the proposed framework to wind turbine dataset

The results regarding the modeling of wind turbine data are presented in Fig. 4.7. In panel (a), the deterministic trend component is highlighted . In particular, this component exhibits dynamic variability over time, aligning closely with our theoretical postulations.

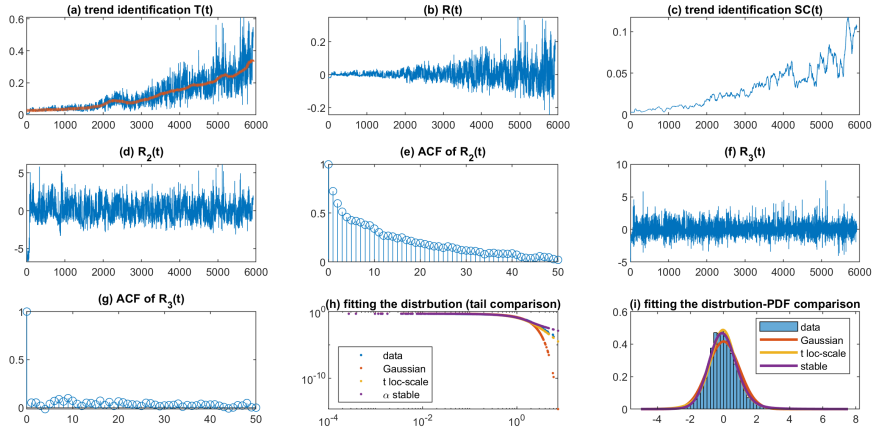


FIGURE 4.7: Wind turbine data modelling - the robust approach. Results of each step of the proposed framework: (a) identified trend $T(t)$, (b) random component $\{R(t)\}$, (c) identified scale $SC(t)$, (d) normalized random component $\{R_2(t)\}$, (e) autocorrelation of normalized random component, (f) residuals of AR model, (g) autocorrelation of the residual series, (h) comparison of fitted tails, (i) comparison of fitted distributions PDF.

Upon eliminating the deterministic trend, the random component is gotten, as illustrated in panel (b). The observed sequance exhibits non-stationary behavior, attributed to its evolving scale over time. Concurrently, the component corresponding to the sequence $SC(t)$ is demonstrated in panel (c), revealing distinct evidence of seasonality within the random component. Furthermore, our developed model is validated within this context, showcasing discernible patterns of growth within each regime ($0-2000$ and $2000-4000$), subsequently evolving into a more complex structure.

The normalized data are presented in panel (d), highlighting a notable dependence within this domain. Then, panel (e) shows the empirical robust ACF, which exhibits significant differences from the FEMTO data. This results in parameters significantly different from zero in the fitted AR model, specifically an AR model of order 4 (AR(4)), with the following estimated parameters: $\phi_1 = 0.54$, $\phi_2 = 0.1$, $\phi_3 = 0.036$, and $\phi_4 = 0.016$. The residual series for the sequence $\{R3(t)\}$ is presented in panel (f), while the empirical ACF for this series is shown in panel (g), indicating a series of seemingly independent observations post-implementation of the inverse AR filter, markedly reducing the data's dependence.

Comparative analysis of empirical and theoretical tails, corresponding to three tested distributions with parameters estimated from the data, suggests a closer resemblance of the residual series to the t location-scale distribution (with $\nu = 6.33$), see Appendix A, over the α-stable distribution, decisively rejecting the Gaussian distribution. This assertion is further supported by contrasting the actual and theoretical PDFs, as depicted in panel (i). The notable attributes of the suitable elements are summarized in Table 4.3.

TABLE 4.3: Identified characteristics for the wind turbine dataset.

Trend	time-varying
Scale	time-varying
Autodependence of random component	significant
Coefficients of AR model	significantly non-zero
Distribution of the random component	t location-scale with $\nu = 6.33$

4.4.3 Discussion

The findings derived from the real datasets confirm the initial hypotheses underlying the respective models. Our analysis discerns time-varying deterministic trends and scale functions present within both datasets. Moreover, autodependence is identified in both cases. It is important to mention that although the autodependence in the FEMTO dataset may not be significant, there is a noticeable autodependence in the wind turbine dataset, highlighting its importance. Both situations exhibit a non-Gaussian distribution in the discovered random component's distribution. The distribution of the first dataset, known as FEMTO data, is more similar to a Gaussian distribution due to its higher ν parameter. In contrast, the wind turbine dataset exhibits impulsive (non-Gaussian) behavior, as indicated by the smaller value of the parameter responsible for this behavior in the fitted t location-scale distribution. In this chapter, the models corresponding to the three regimes outlined in Section 3.2 as separate entities is not analyzed. Here, it is used the simplistic premise that the random component of the analyzed datasets is represented by a single stochastic model. Furthermore, it is assumed that the distribution of this model remains constant throughout time.

In order to validate the findings, the quantile lines produced by the fitted models when applied to the two real datasets is carefully examined. It is important to note that the display of quantile lines obtained from the fitted model is a common way to determine how well the model fits the data. This approach is widely used in both academic literature and real-world applications. The process of constructing the quantile lines involves several steps. Firstly, the model to actual data and obtain a fit is applied. Next, a certain number of trajectories based on the fitted model is simulated. Lastly, for each time point, the quantiles at specific levels, creating is computed what it is call quantile lines. If the real data falls within these calculated intervals as expected, it indicates that the fitted model is accurate. These quantile lines, derived from various fitted models, can also help us figure out the best model or indicate which one among those studied fits the real data the best.

The models have been synthesized and the simulation results are shown in Fig. 4.8. The image displays the genuine datasets using gray lines (FEMTO data in the top panel and wind turbine data in the bottom panel). The generated quantile lines, representing the 5% and 95% levels, are shown in black. These quantile lines were obtained by simulating 100 trajectories based on the fitted models.

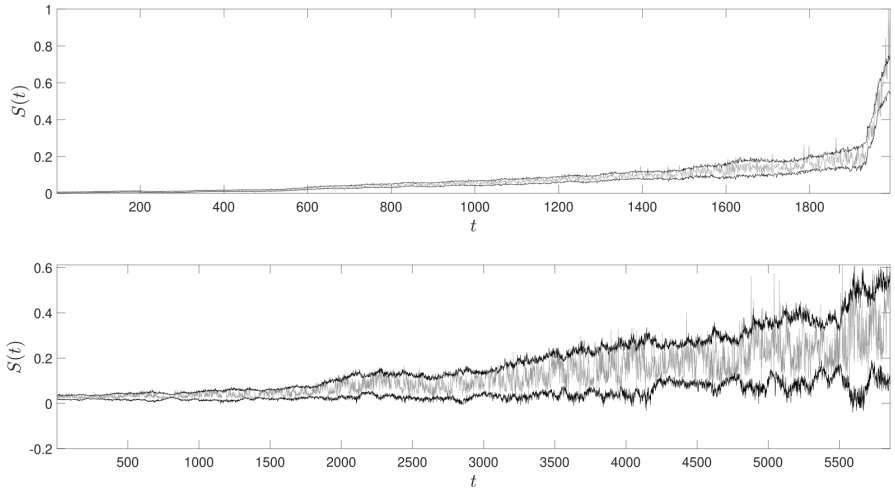


FIGURE 4.8: The validation of the approach for real datasets (represented by grey lines) is demonstrated: the top panel showcases the FEMTO data, while the bottom panel displays the wind turbine data. Moreover, the constructed quantile lines (depicted by black lines) at the 5% and 95% levels are established based on 100 simulated trajectories corresponding to the fitted non-Gaussian models.

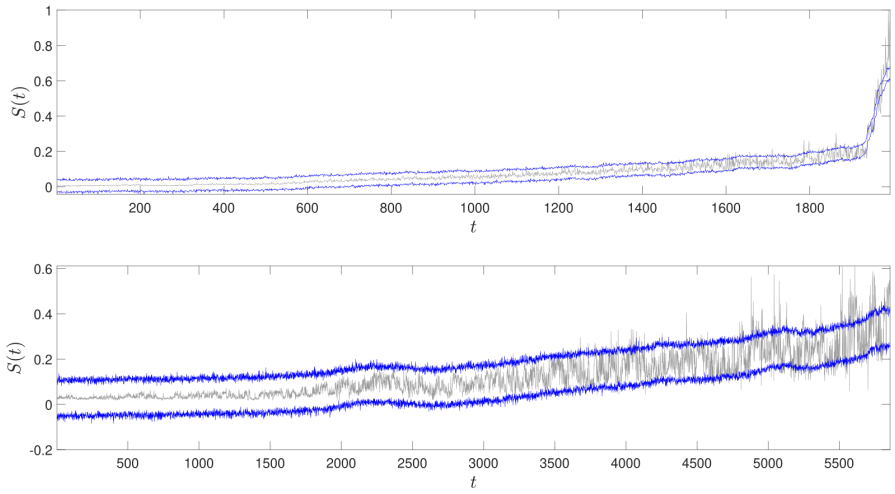


FIGURE 4.9: The validation of the approach for real datasets (depicted by grey lines) is presented: the top panel represents the FEMTO data, while the bottom panel displays the wind turbine data. Additionally, the constructed quantile lines (illustrated by blue lines) at the 5% and 95% levels are based on 100 simulated trajectories corresponding to the fitted Gaussian models.

To confirm the suitability of employing the non-Gaussian model with non- homogeneous characteristics, the quantile lines in Fig. 4.9 are presented. These lines are produced using a simple Gaussian model, which includes a time-varying trend and Gaussian noise. This model represents the standard approach for data modeling in such scenarios. Similarly to before, the quantile lines are created using the 100 trajectories of the fitted model, with the quantile values set at 5% and 95%. In this case,

the moving median approach was used to fit the deterministic trend and the maximum likelihood method was used to estimate the Gaussian distribution parameters for the detrended data. By comparing Figs. 4.8 and 4.9, it becomes evident that the classical model is not suitable for the actual datasets. The Gaussian models seem to overestimate the data from the initial regime (healthy stage) and underestimate the data from the subsequent section (critical stage).

As a result, fitted models do not adequately capture the distinctive features of the datasets and cannot be deemed ideal, especially for prognostic purposes. Conversely, upon analyzing the quantile lines of the proposed non-Gaussian non- homogeneous model, it becomes apparent that the simulated and actual data coincide seamlessly for both datasets.

To demonstrate the superiority of the proposed model over the Gaussian-based technique, Fig. 4.10, it is shown the quantile lines at the 95% level for both datasets produced using both approaches. The quantile lines are represented by black and blue lines, respectively. In addition, the difference between the two approaches is shown. It can be inferred that the quantiles for the first regime, which represents the healthy stage, have greater values in the Gaussian-based model compared to the non-Gaussian-based method. However, in the critical stage, the situation is reversed. The significant differences in quantiles between the non-Gaussian and Gaussian-based models underscore the strong suitability of the proposed model. Ultimately, the RMSE between the quantiles for models based on non-Gaussian and Gaussian distributions is computed. The FEMTO data yield a value of 0.0239, however, the wind turbine data yields a value of 0.0540. The disparity in RMSE values is associated with the variability in the scale function and the non-Gaussian nature of the residuals. The FEMTO data were found to follow a t location-scale distribution with a higher number of degrees of freedom compared to the wind turbine data. This suggests that the FEMTO data exhibit a distribution that is more similar to a Gaussian distribution than the wind turbine data. It is vital to note that the selection of window size, done by trial and error, for trend estimate and scale identification is significant but not crucial. The window size should be large enough to minimize bias in the generated statistics, while also small enough to ensure precision in the time domain. The window size for the trend and the scale differs because the rate of change for the trend is slower than the rate of change for scale. The size of the window does not have a direct impact on the calculation of the ACF and the applicable AR model.

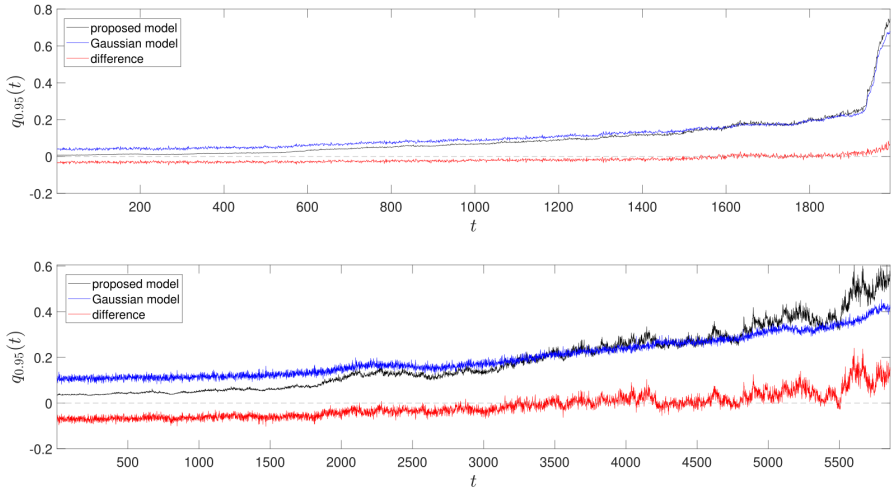


FIGURE 4.10: The quantile lines at the 95% level for the FEMTO data (top panel) and wind turbine data (bottom panel) are compared for two models: the suggested non-Gaussian non-homogeneous model and the Gaussian-based model, which includes Gaussian noise and a deterministic trend. The red lines represent the differences between the computed quantile lines. The quantile lines were produced using 100 trajectories of the simulated models in all situations.

4.5 Chapter highlights

This chapter addresses the challenge of identifying characteristics in long-term health index data from a condition monitoring system. A new framework introduces new components to the classical approach, particularly suited for non-Gaussian noise distributions and time-varying random characteristics. The efficacy of proposed approach is evaluated throw the simulated health index in the presence of different levels of non-Gaussian noise. The proposed framework is also applied on two real datasets. In both real cases the results confirm the time-varying characteristic component in the deterministic and random parts. However, the results has been demonstrated that the distribution of the noise in case of FEMTO dataset is close to the Gaussian distribution, however in case of the wind turbine the distribution of noise is far from the Gaussian distribution.

The novel insights and framework presented in this chapter allow for the creation of advanced models that closely mimic real degradation trajectories, particularly in case of the presence of the non-Gaussian noise. These simulated data can be used to train machine learning-driven prognostic systems.

Furthermore, the tools and framework introduced enable effective segmentation of health index data for advanced modeling, such as predicting the RUL. This segmentation improves predictive maintenance strategies by facilitating the targeted analysis of different degradation stages. The methodologies discussed help practitioners understand the underlying patterns and behaviors, leading to more reliable RUL predictions and improved decision making in asset management and maintenance planning.

Chapter 5

Segmentation of health index data

This chapter primarily focuses on the segmentation of the HI. Segmenting the HI provides an opportunity to construct precise models for future analysis, such as predicting RUL. However, due to the complex behavior of the HI, including non-Gaussian noise, this segmentation task necessitates the use of appropriate tools, as discussed in previous sections.

Therefore, this chapter is divided into two main subsections. The first subsection investigates the segmentation problem based on historical HI data, while the second addresses the problem as a one-step prediction task. This division allows for a comprehensive exploration of segmentation methodologies that cater to different analytical scenarios and data processing requirements.

5.1 Segmentation of long-term historical health index data

In this subsection, the segmentation problem is addressed using historical data. The degradation process is assumed to consist of three stages, each characterized by a specific deterministic trend. Based on this understanding, a segmentation method is developed to effectively identify and delineate these distinct stages within the degradation process.

5.1.1 Methodology of segmenting long-term historical health index data

The long-term HI data is presumed to have three distinct stages with different behaviors. The goal is to divide the time series into three parts using robust regression methods to achieve the most accurate fit for each model.

5.1.2 Piece-wise regression models

This category of techniques enables the data to be divided into three pieces, each characterized by a distinct model form. Three functional components are utilized: constant, linear, and exponential, as illustrated below

$$\widehat{y}_t(\Theta) = \begin{cases} f_1(t,\theta_1) = \theta_1, & 0 < t \leq \tau_1, \\ f_2(t,\theta_2,\theta_3) = \theta_2 t + \theta_3, & \tau_1 < t \leq \tau_2, \\ f_3(t,\theta_4,\theta_5,\theta_6) = \theta_4 \exp(\theta_5 t) + \theta_6, & \tau_2 < t \leq N, \end{cases} \tag{5.1}$$

here, $\widehat{y}_t$ represents the estimation of y, which is the observed series of HI. The functions $f_1(.)$, $f_2(.)$, and $f_3(.)$ denote constant, linear, and exponential functions, respectively. In addition, the variables τ_1 and τ_2 represent changing point between healthy stage and degradation stage (CP1) and changing point between degradation stage to critical stage (CP2), respectively. The quantity N corresponds to the length of the

signal, while $\theta_1, ..., \theta_6$ are constants that need to be fitted. These constants are used to simulate the deterministic components of HI.

Ordinary least squares

The ordinary least squares (OLS) is fundamental technique relies on reducing the sum of the squares of the residuals of the model. The procedure designed to find optimal set of parameters iterates in two loops through possible divisions into three windows, allowing identification of the optimal values for τ_1 and τ_2

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{t=1}^{\tau_1-1} (y_t - f_1(t, \theta_1))^2 + \sum_{t=\tau_1}^{\tau_2-1} (y_t - f_2(t, \theta_2, \theta_3))^2 + \sum_{t=\tau_2}^{N} (y_t - f_3(t, \theta_4, \theta_5, \theta_6))^2. \tag{5.2}$$

When dealing with observations that have Gaussian noise, the best parameters can be calculated using MLE using residuals that are distributed according to a Gaussian distribution.

Dynamic programming segmentation

This approach utilizes optimized dynamic programming to achieve optimal segmentation of the time series of observations $y_t = \{y_1, y_2, \cdots, y_N\}$. The model parameters are estimated on the basis of the MLE using a Gaussian distribution. The computational expense of this paradigm is substantial when using dynamic programming. To obtain further information about this approach, please refer to the following sources [131, 132].

Iterative reweighted least squares

The iterative reweighted least squares (IRLS) method is a modified version of the OLS method. In IRLS, weights are assigned to error terms based on their changing variance, which is assumed to be finite. For each time t, the weight w_t is equal to the reciprocal of the corresponding standard deviation σ_t of the observation error term. The weights are adjusted iteratively. In each step, denoted by j, the vector of weights $w_t^{(j)}, t = 1, \ldots, N$ is used to update the parameters of the piecewise regression

$$\widehat{\Theta}^{(j)} = \underset{\Theta}{\operatorname{argmin}} \sum_{t=1}^{\tau_1-1} w_t^{(j)} (y_t - f_1(t, \theta_1^{(j-1)}))^2 + \sum_{t=\tau_1}^{\tau_2-1} w_t^{(j)} (y_t - f_2(t, \theta_2^{(j-1)}, \theta_3^{(j-1)}))^2 + \sum_{t=\tau_2}^{N} w_t^{(j)} (y_t - f_3(t, \theta_4^{(j-1)}, \theta_5^{(j-1)}, \theta_6^{(j-1)}))^2. \tag{5.3}$$

The variable $\widehat{\Theta}^{(j)}$ represents the current set of values for Θ in the j-th iteration of the IRLS algorithm. One of the iterative approaches used to determine appropriate weights is the Tukey biweight function $\phi_{Tu}(y) = y(1 - y^2)^2$, which is applied to the residuals of the model. This method is implemented in the Matlab `robustfit()` function, as stated in [133]. The weights $w_t^{(j)}$ are adjusted using the formula described in [134].

Least absolute error

The least absolute error (LAE) utilized the cost function which is the sum of absolute deviations, rather than the sum of squares. The method described in [135] is a well-established alternative to the OLS method. It is particularly useful for making robust adjustments to regression parameters, even when dealing with infinite variance. The optimization technique, which incorporates the LAE cost function with varying points, is specified as follows

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{t=1}^{\tau_1-1} |y_t - f_1(t,\theta_1)| + \sum_{t=\tau_1}^{\tau_2-1} |y_t - f_2(t,\theta_2,\theta_3)| + \sum_{t=\tau_2}^{N} |y_t - f_3(t,\theta_4,\theta_5,\theta_6)|. \tag{5.4}$$

Similarly to the OLS approach, all potential divisions are evaluated in three stages to determine τ_1 and τ_2. This approach is advantageous when the errors exhibit a heavy-tailed distribution rather than a Gaussian distribution.

5.1.3 Student's t distribution estimation

Student's t distribution estimation (ST) method assumes that the residuals of the model in each degradation stage follow a scaled Student's t distribution

$$r_t = \begin{cases} \sigma_1 u_t, & 0 < t \leq \tau_1, \\ \sigma_2 u_t, & \tau_1 < t \leq \tau_2, \\ \sigma_3 u_t, & \tau_2 < t \leq N. \end{cases} \tag{5.5}$$

The series $u_t = \{u_1, u_2, \cdots, u_N\}$ consists of independent random variables that follow a Student's t distribution. Each segment of the series, denoted by $i = 1, 2, 3$, has ν_i degrees of freedom. This feature allows for the presence of extreme values in the error distribution and enhances the ability to accurately model non-Gaussian observation data by providing a more resilient fitting process. The estimation approach relies on the MLE technique, which is applied to every potential split into three stages. The optimal collection of parameters is determined to satisfy the following criteria

$$\begin{aligned} \widehat{\Theta} = & \underset{\Theta}{\operatorname{argmin}} \sum_{t=1}^{\tau_1-1} \log\left(p_{\nu_1}\left(\frac{y_t - f_1(t,\theta_1)}{\sigma_1}\right)\right) \\ & + \sum_{t=\tau_1}^{\tau_2-1} \log\left(p_{\nu_2}\left(\frac{y_t - f_2(t,\theta_2,\theta_3)}{\sigma_2}\right)\right) \\ & + \sum_{t=\tau_2}^{N} \log\left(p_{\nu_3}\left(\frac{y_t - f_3(t,\theta_4,\theta_5,\theta_6)}{\sigma_3}\right)\right), \end{aligned} \tag{5.6}$$

where $p_\nu(\cdot)$ is the PDF of Student's t distribution , see Appendix.A.

Furthermore, it is important to mention that the most favorable values for the parameters of the methods explained in this section (namely OLS, IRLS, LAE, and ST) are discovered by unconstrained multivariate optimization using a derivative-free approach. The `fminsearch` function in Matlab is utilized for this purpose.

5.1.4 Hidden Markov models and Expectation-Maximization algorithm

For a shorter notation, in this subsection, the process of observations is denoted as $y_t = \{y_1, y_2, \cdots, y_N\}$. . HMM are the models in which to describe the evolution of the observation process X, an additional unobserved process (hidden states) $z_t = \{z_1, z_2, \cdots, z_n\}$ is introduced so that the whole system has the Markov property. As hidden states are unknown, the full likelihood of a system $P(y, z|\Theta)$ cannot be calculated directly to obtain the parameters using the MLE method. The standard iterative method that can be used instead is called Expectation-Maximization (EM) algorithm. First, some initial values of $\Theta^{(0)}$ are set. The next EM procedure is based on repeated steps in calculating the likelihood expectation given the distribution of z conditioned with y and $\Theta = \Theta^{(u)}$ and reassigning Θ to new values that maximize the conditional expectation. The u-th step can be described with the update formula

$$\Theta^{(u+1)} = \underset{\Theta}{\operatorname{argmax}}\, E_{z|y,\Theta=\Theta^{(u)}}\Big(P(y, z)|\Theta\Big), \tag{5.7}$$

where the letter E stands for conditional expectation. Two different models from this class are considered. However, they share the same equation for evolution of X given z, which is a polynomial (of degree p) regression

$$y_t = \sum_{i=0}^{p} \beta_{i,z_k} t^i + \sigma_{z_t} \epsilon_t, \tag{5.8}$$

where $t = 1, \ldots N$, the constants β_{0,z_k} to β_{p,z_k} are polynomial coefficients in the state (degradation stage) z_k, and ϵ_t is Gaussian white noise. The set of $p + 1$ polynomial coefficients depends on the current stage z_t. The hidden states can be interpreted as degradation stages indexed from 1 to K in the proper order from healthy state to full degradation (thus K denotes the number of distinct degradation stages). In this thesis, according to the framework described in Section 2, K is set to 3.

In the HMM approach, CP1 and CP2 are not included in Θ. The result of the fit of the HMM model to the data contains probabilities for each t of the current stage z_t equal to $1, \ldots, K$. The state with the highest probability indicates the degradation state of the observed machine at time t, thus obtaining the division of the degradation process into three stages.

Hidden Markov model regression

Hidden Markov model regression (HMMR) approach employs a mixture of polynomial regressions handled by hidden states of a discrete-time Markov chain. Changing of the stages is therefore fully described by the initial distribution and the one-step transition matrix. Significant restrictions are therefore placed on the transition probabilities to better reflect the nature of the process. Thus, for all $l = 1, \ldots, K$, the following is obtained

$$P\Big(z_{t+1} = k|z_t = l\Big) = 0, \quad \text{for } k \notin \{l, l+1\}, \tag{5.9}$$

and generally nonzero otherwise. The parameters are estimated through the Baum-Welch algorithm as described in [131].

Hidden logistic process

In Hidden logistic process (HLP) method, the unobserved switching stage process affecting the set of polynomial regression coefficients is modeled with a multinomial logistic regression model. The distribution of the current stage z_t in the HLP is assumed to be as follows

$$\pi_{t,k}(\beta) = P(z_t = k|\beta) = \frac{exp(\sum_{i=0}^{p} \beta_{ki} t^i)}{\sum_{l=1}^{K} exp(\sum_{i=0}^{p} \beta_{li} t^i)}, \qquad \text{for } k = 1, \ldots, K, \tag{5.10}$$

where $\beta = [\beta_{ki}]_{k=1,\ldots,K,\, i=0,\ldots,p}$ is matrix containing $p+1$ regression parameters for all K degradation stages. As z is not observed, the Gaussian distribution of y_t given z_t must be included in the total log-likelihood of the model. In terms of probability density functions, the following is obtained (see [131])

$$p(X|\Theta) = \sum_{t=1}^{N} \log \sum_{k=1}^{K} \pi_{t,k}(\beta) p_{\Theta}(y_t|z_t = k), \tag{5.11}$$

where the Gaussian PDF denoted with $p_{\Theta}(y_t|z_t = k)$ comes from Eq. (5.8). The resulting model is a special case of a time-heterogeneous Gaussian mixture model. A dedicated EM for this approach is described in [131]. As shown there, in a single iteration of EM the expectation term is a function of Θ (compare Eq. (5.7)) can be separated into two terms, one dependent on β and the other dependent on σ and β. Thus, fitting of these two sets of parameters can be done independently. Matrix β is estimated utilizing a multi-class IRLS approach.

5.1.5 Analysis of simulated data

Based on the assumption of Section 3.2, the long-term data model is used to generate HI observations. After the mentioned methodologies are employed to segment the data into three stages in the presence of Gaussian noise and non-Gaussian noise. In addition, to show the performance of the methodology, the results were compared together.

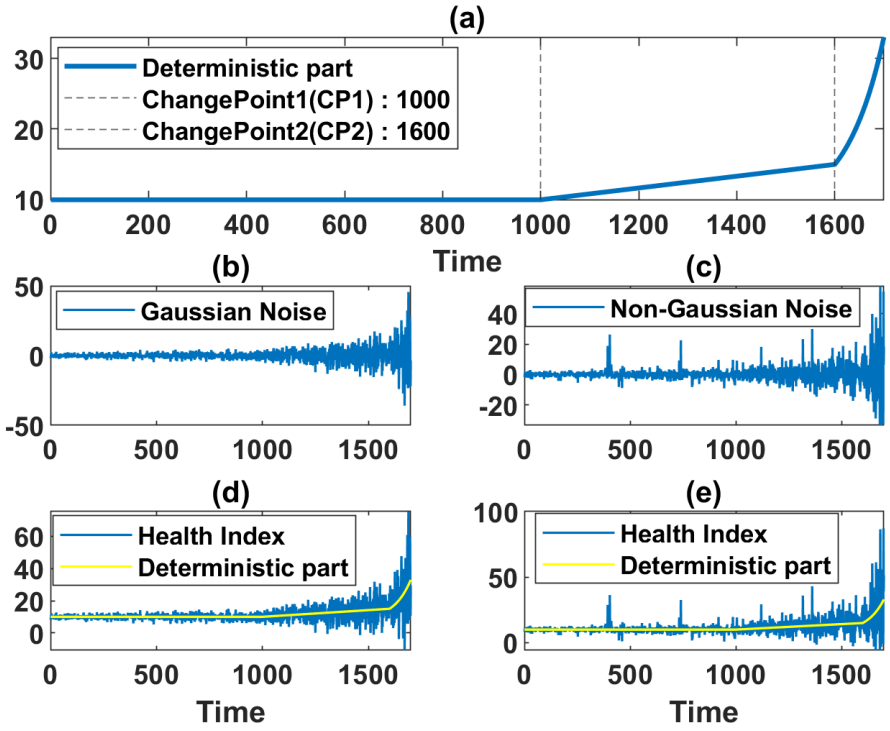


FIGURE 5.1: Simulated HI: (a) deterministic part with marked changing points, (b) Gaussian noise, (c) non-Gaussian noise with $\nu = 3$, (d) HI in presence of Gaussian noise, (e) HI in presence of non-Gaussian noise.

Analysis for Gaussian distributed model

The proposed methodology is implemented in this subsection using the data generated by the proposed model, assuming that the noise term follows a Gaussian distribution with mean 0 and standard deviation 1, denoted as $\widetilde{R}(t) \sim \mathcal{N}(0,1)$. The segmentation results for the simulated data are collectively shown for all of the chosen methods. As a result of the simulation process, the healthy stage will transition to the degradation stage (CP1, represented by τ_1) at Time=1000, and CP2 (represented by τ_2) will occur at Time=1600.

The exemplary simulated trajectory, along with the predicted changing points CP1 and CP2 resulting from each of the methods, can be shown in Fig. 5.2. It is evident that in this scenario, when Gaussian noise is present in all stages, the chosen algorithms were able to accurately detect CP1. CP1 is accurately detected by most approaches, including ST, LAE, and IRLS. Other methods also identify this moment around between Time=1000 and Time=1088, which is acceptable. However, the HLP method detects this point at Time=882, earlier than the required time. On the basis of this fact, the transition between the last two phases is noteworthy. CP2 was identified using most approaches, except for IRLS, OLS and HLP, which were unable to identify this point as suitable.

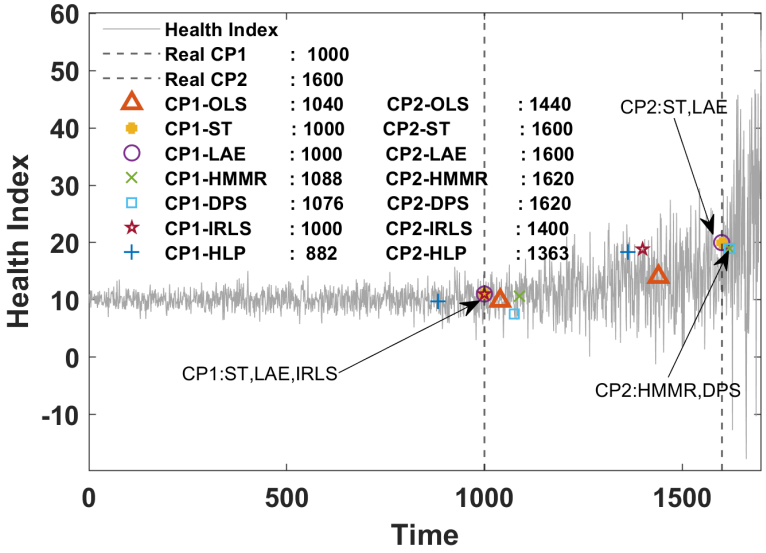


FIGURE 5.2: Changing points detection in the presence of Gaussian noise.

The estimation technique was iterated 100 times using the same model. The estimation results are depicted in boxplot in Fig. 5.3. Horizontal red lines correspond to the mean estimation results, while the gray dashed line represents the real value of the change point. Estimation of CP1 and CP2, determined by HMMR, HLP, and DPS, exhibits minimal variation, as indicated by the tiny value of the IQR. However, in the case of CP1, the result is positively skewed away from the true value of 1000. Regarding CP2, most of the methods (except IRLS, OLS and ST) exhibit a notable presence of outliers on the lower end, deviating from the actual value of 1600.

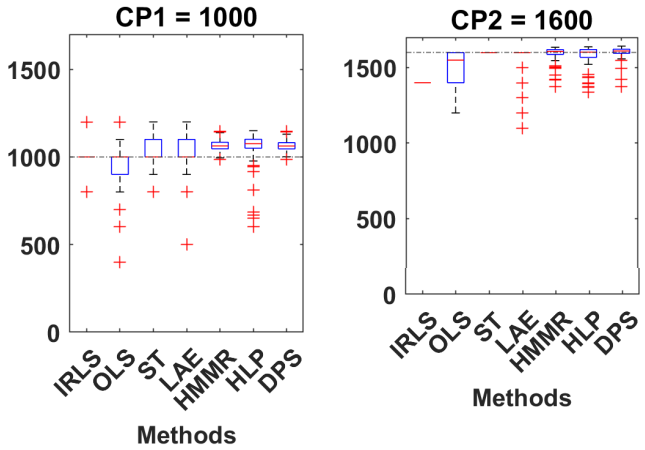


FIGURE 5.3: Monte Carlo analysis for changing point detection in presence of Gaussian noise, left subplot is boxplot of detected CP1 and right subplot is boxplot of detected CP2.

Additionally, the mean squared error (MSE) is computed by comparing the genuine changing point values (CP1=1000 and CP2=1600) with the estimations obtained from the Monte Carlo simulation conducted for each approach. The illustration can be seen in Fig. 5.4. As depicted in Fig. 5.4, the IRLS approach exhibits the lowest MSE while identifying CP1, but the greatest MSE when detecting CP2. In addition,

the ST technique has a high capability to accurately detect CP2. Nevertheless, it has failed to match CP2's performance in detecting CP1. Upon doing a comprehensive comparison of all the bars, it can be deduced that the DPS, HMMR, and ST techniques exhibit the lowest MSE when subjected to Gaussian noise.

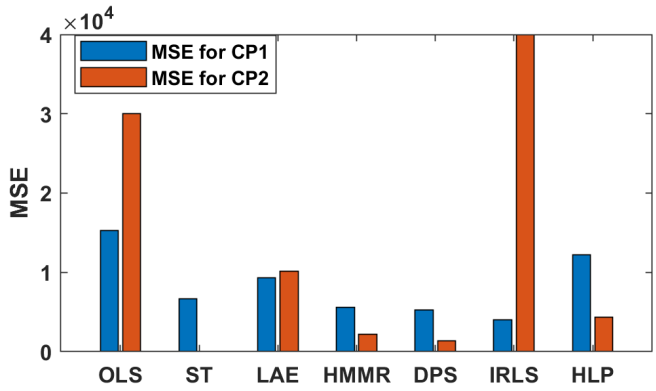


FIGURE 5.4: MSE for Monte Carlo analysis for changing point detection in presence of Gaussian noise.

Analysis for non-Gaussian distributed model

In this section, the suggested approach is used to analyze the data produced by the suggested model with a random component that follows a Student's t distribution. The analysis considers various values of the parameter ν, ranging from just above 2 to a sufficiently large value. The former represents the limit case where the variance becomes infinite, resulting in a highly non-Gaussian distribution. The latter represents a value where the distribution closely resembles a Gaussian distribution.

Based on the simulation technique, CP1 and CP2 have values of Time=1000 and Time=1600, respectively. Fig.5.5 displays a simulated trajectory, along with the estimated changing points CP1 and CP2, obtained from different approaches. This example refers to a scenario with degrees of freedom $\nu = 2.1$. CP1 is accurately identified using methods such as ST and LAE, whereas other approaches do not detect this point correctly. For example, HMMR, DPS, IRLS, and HLP identify this point at Time=347, 348, 600, and 900, respectively. Additionally, OLS detects this point with a delay at Time=1280. ST and LAE accurately detected CP2, while HMMR and DPS detected it with a small delay at Time = 1612, which is acceptable. However, IRLS on OLS identified this point at Time=1400 and 1440, which is significantly earlier than expected. The results of HLP are deemed insignificant.

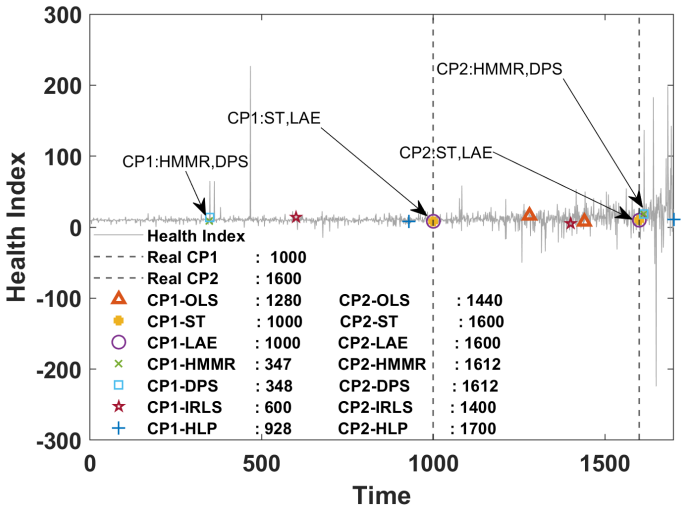


FIGURE 5.5: Segmentation results for non-Gaussian simulation case with $\nu = 2.1$.

In order to validate the effectiveness of the techniques, the estimation procedure was replicated 100 times using the same model but with varying degrees of non-Gaussian noise. The estimation results are depicted using boxplot in Fig. 5.6. Upon evaluating the box plot, it is evident that the ST approach outperforms other methods in detecting changing points, particularly CP2, in the presence of varying levels of non-Gaussian noise.

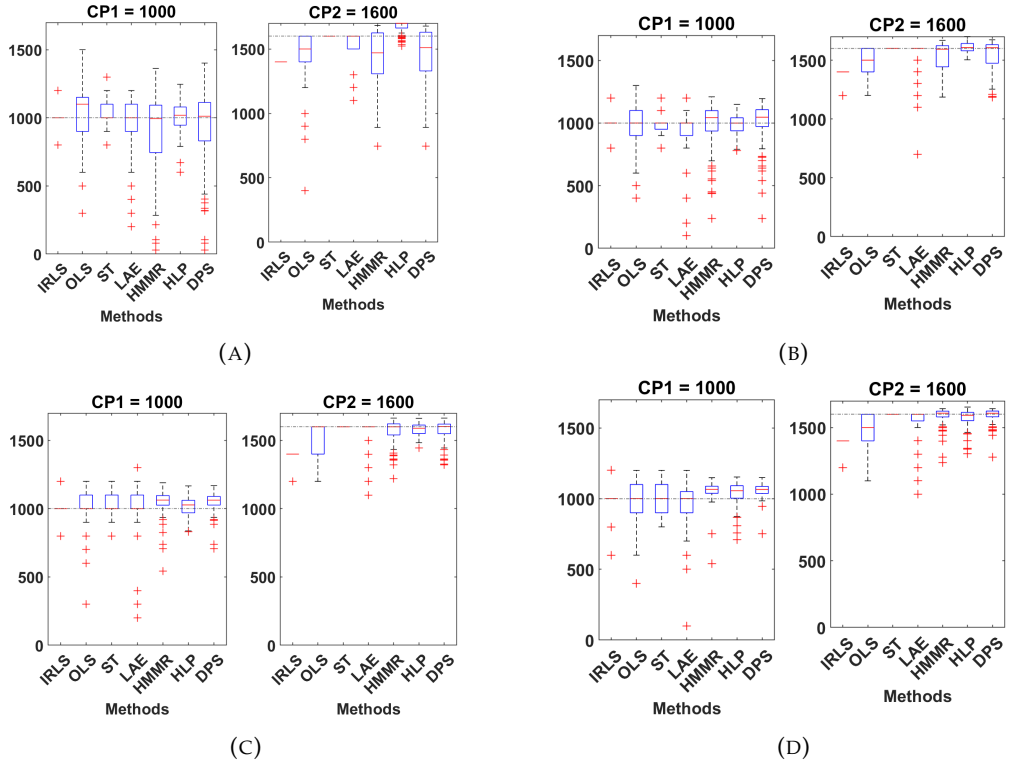


FIGURE 5.6: Monte Carlo analysis for changing point detection in the presence of different levels of non-Gaussian noise, (a) $\nu = 2.1$, (b) $\nu = 3$, (c) $\nu = 5$, (d) $\nu = 10$.

Furthermore, the MSE is computed using actual changing points (CP1=1000 and CP2=1600) for each Monte Carlo simulation, considering various values of ν. The resulting MSE values are illustrated in Fig. 5.7. Panel (a) of Fig. 5.7 illustrates that the MSE of several methods used to detect CP1, including OLS, HMMR, DPS, and to some extent HLP, has shown an increase as ν decreases. This decrease in ν leads to an increase in impulsiveness. However, decreasing the ν value does not significantly impact the MSE of robust methods such as ST, LAE, and IRLS. Furthermore, it should be noted that ST, IRLS and HLP methods exhibit the lowest MSE when it comes to detecting CP1, outperforming all other specified approaches. Similarly, the MSE of the approaches for CP2 is displayed in panel (b) of Fig. 5.7. The MSE of methods such as HMMR, DPS, and OLS increased as ν decreased, while the MSE of methods like ST, LAE, IRLS, and HLP remained unaffected by a decrease in ν. In addition, the ST approach has the ability to detect CP2 with the lowest MSE compared to all other methods.

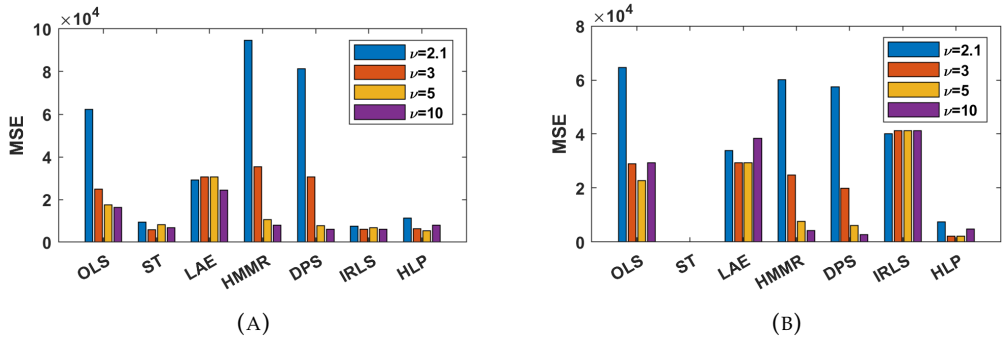


FIGURE 5.7: MSE for Monte Carlo analysis for changing point detection in the presence of different levels of non-Gaussian noise, (a) bar plot of MSE for CP1, (b) bar plot of MSE for CP2.

To establish the universality of the suggested strategy, the methodology was also validated for a different non-Gaussian distribution, specifically the α-stable distribution [136]. Random variables following a α-stable distribution, excluding the Gaussian case, exhibit infinite variance. Consequently, the task of identifying changing points becomes far more challenging compared to the scenario discussed in this study, even for the non-Gaussian distribution under analysis.

5.1.6 Real data analysis

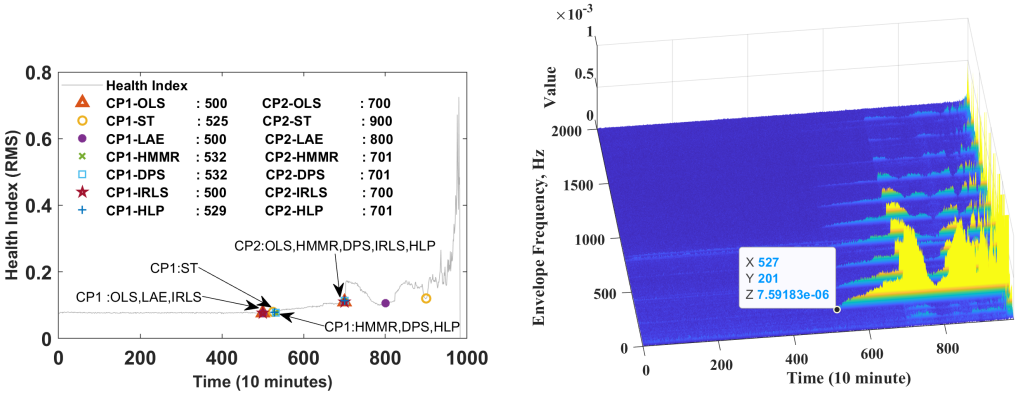
In this part, the suggested methodology will be implemented and evaluated using existing real datasets to evaluate the performance of proposed approach in real case.

Results for IMS dataset

The segmentation methods' results for this case study are shown in panel (a) of Fig. 5.8. Panel (a) of Fig. 5.8 clearly shows that most approaches identified the first change point (CP1), which marks the transition from the healthy stage to the degradation stage, occurring between Time=500 and 560. Based on a visual examination of the HI and the corresponding outcomes, it appears that most of the approaches produced satisfactory results. For the second change point (the point where the trend shifts from the degradation stage to the critical stage), many approaches including OLS, HMMR, DPS, IRLS, and HLP identified this point roughly at Time = 700, where there was a significant increase in the value of HI. The LAE approach has discovered

at Time=800, and subsequently, the HI values exhibit rapid growth. Unfortunately, both of the stated occurrences appear to be inaccurate representations of CP2 points on a global scale, as the HI values have been shown to decrease. It is claimed that CP2 = 900, as determined by the ST technique, is accurate due to the trend of exponential decline observed in the data beyond that point, which spans a significant period of time. To validate these findings, envelope analysis is employed on each set of vibrations and the results are subsequently plotted in panel (b) of Fig. 5.8. For further information on envelope analysis, the following references are recommended: [137, 138, 139, 140].

In panel (b) of Fig. 5.8, it is evident that after Time = 527, a harmonic frequency has emerged, specifically the ball pass frequency outer (BPFO) fundamental harmonic. This indicates that the bearing has entered the degradation stage. After comparing the envelope analysis and segmentation approaches, it can be inferred that the ST and HLP methods yield the most optimal result.



(A) (B)

FIGURE 5.8: Changing points detection IMS datasets case number 2, (a) result of changing points detection method, (b) envelope spectrum of the raw vibration signal.

Results for FEMTO dataset

The results of the segmentation methods for this case study are displayed in Fig. 5.9. It is evident that this dataset adheres precisely to the concept of three distinct stages: from Time = 0 to around 1300, it remains relatively constant, followed by a linear increase up to Time = 2700, and finally experiencing fast development. There is noticeable noise, particularly in the middle stage, and some minor outliers can also be observed. Therefore, it is classified as a pattern that exhibits almost Gaussian noise. As depicted in Fig. 5.9, most algorithms identified the initial change point (the point where the transition from a healthy to a degraded state occurs) between Time=1200 and 1400, with the exception of HMM-based algorithms such as HMMR and HLP, which detected the first change point significantly later than it actually occurred. According to the initial visual inspection, the first point of change should be identified around Time = 1000. This indicates that the ST approach has the ability to identify this spot with a tolerable margin of error. It is important to note that the transition between healthy and degraded stages is gradual and not easily detectable. Based on the information provided in Fig. 5.9, the second changing point is expected to occur approximately at Time=2700. At this moment, the HI is observed to undergo a significant and rapid increase on a worldwide scale. Thus, it can be inferred that the

ST, OLS, HMMR, and HLP algorithms are capable of detecting this point. However, the LAE, DPS, and IRLS approaches have failed to accurately identify this spot.

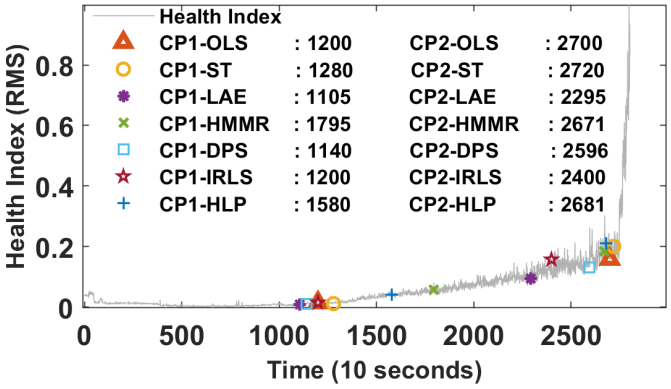


FIGURE 5.9: Result of changing points detection methods for the FEMTO datasets.

Results for wind turbine dataset

In order to implement the suggested method on the wind turbine datasets, the inner race energy is chosen as the health indicator. Fig. 2.2 clearly displays a significant number of outliers in this dataset. Therefore, it is classified as a pattern exhibiting non-Gaussian noise. Furthermore, certain oscillations have been observed in the HI, which have been found to be associated with changes in the load or certain phenomena such as self-healing. This particular behavior poses considerable difficulty for the segmentation and prediction of RUL. The results of the segmentation methods for this case study are provided in Fig. 5.10. The CP1 (change point indicating the transition from a healthy stage to a degrading stage) is identified by all methods between the dates of 15 March and 21 March. Specifically, OLS, ST, LAE and IRLS detected this point on 15 March, while HMMR, DPS, and HLP identified it on 18 March, 18 March and 19 March, respectively.

Based on the analysis of Fig. 5.10, it appears that 15 March is a more suitable choice for CP1. This is because, following this date, the HI experiences a significant increase until 21 April. This suggests that the point in question is detected by HMMR, HLP, and DPS with a certain delay. On 8 April, a CP2 indicating the transition from the deterioration stage to the critical stage trend was identified for HMMR, DPS, and HLP. Regrettably, all the stated approaches appear to incorrectly identify CP2 locations on a global scale. Furthermore, it is incorrect for LAE to have identified CP2 when HI values began to decline. Meanwhile, ST and IRLS detected it on April 20th. Based on the trend of the data showing an exponential development with reduced fluctuations over a significant period of time, it can be inferred that CP2=20 April is the accurate CP2. A separate iteration of these datasets is used to verify the precision of the findings. This version of the datasets consists of a raw vibration signal that is captured for a duration of six seconds every day. The sampling rate for this signal is 100k. You can refer to panel (a) in Fig. 5.11 for a visual representation. Therefore, it signifies that it has an exceptional potential to implement diverse frequency analysis in order to perform condition monitoring. The envelope analysis is performed on each set of vibrations, and the resulting data are progressively presented in panel (b) of Fig. 5.11. Panel (b) of Fig. 5.11 clearly shows the emergence of the harmonic frequency, specifically the fundamental harmonic of the

ball pass frequency , after 14 March. This indicates that the bearing has entered a stage of degeneration. Furthermore, following the date of 22 April, there was a significant and sudden increase in the amplitude of the harmonic frequency, indicating a potential occurrence of CP2. Upon analyzing the results of the segmentation approach, it is evident that the IRLS and ST methods exhibit the most favorable results for this particular scenario.

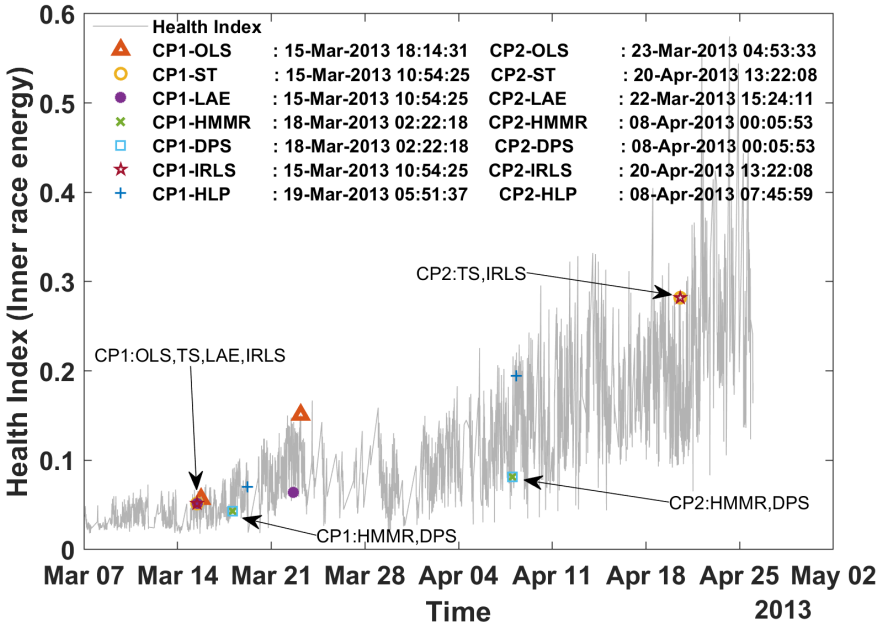
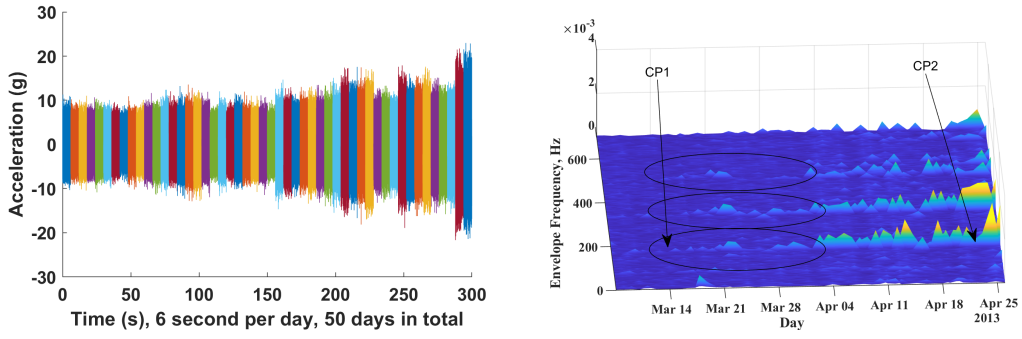


FIGURE 5.10: Changing points detection of wind turbine datasets.



(A) (B)

FIGURE 5.11: Changing points detection wind turbine, (a) raw vibration data, (b) envelope spectrum map.

Discussion

The percentage error for detecting the changes in all the actual datasets is demonstrated in Fig. 5.12. As can be seen in both sub-figures for both CP1 and CP2, the ST method has the lowest percentage error. In contrast, error percentage for the rest of the methods has fluctuated on different datasets, which can include the fact that considering non-Gaussian noise distribution improves the process of detecting changing points in long-term condition monitoring data.

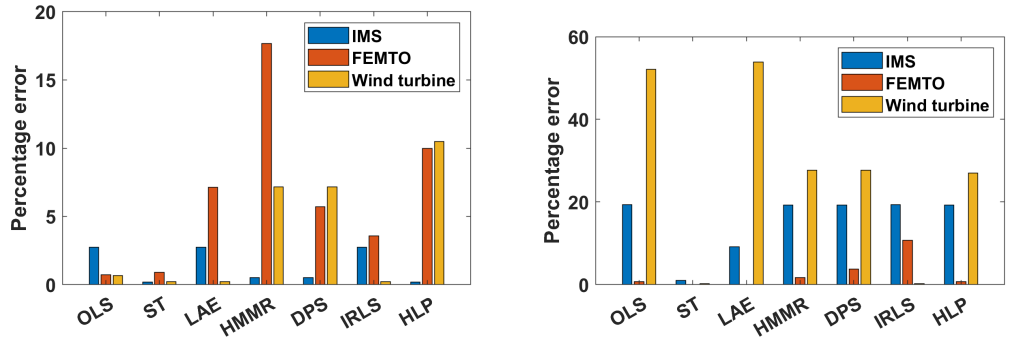


FIGURE 5.12: Percentage error for real datasets changing points detection, (a) percentage error for CP1, (b) percentage error for CP2.

5.2 Probabilistic segmentation of long-term health index data

Previously, the possibility of dividing long-term historical data into three stages using the deterministic component was showcased. In addition, the influence of non-Gaussian noise on the effectiveness of segmentation techniques was clarified.

However, real-life scenarios require an online method for practical functionality. Henceforth, in the subsequent part of this section, the process of segmenting such data using an online technique will be demonstrated, thereby enhancing its applicability in real-world scenarios. The following introduces three dynamic models to characterize different stages of machine health: the healthy stage (constant trend), the degradation stage (linear growth) and the critical stage (exponential or polynomial behavior). These models facilitate the investigation of the degradation process. A robust Bayesian approach has been developed for these dynamic models to perform one-step state predictions. This method is integrated with a dynamic linear model approach to construct a switching model, which provides both one-step predictions of the HI sequence and probabilistic results for the online segmentation of HI data simultaneously.

5.2.1 Methodology of probabilistic segmentation of long-term health index data

Bayesian approach is a highly promising strategy for online segmentation. Significantly, it provides an opportunity to utilize the information from historical data acquired in Chapter 4 for the purpose of segmentation. Furthermore, the Bayesian technique offers the benefit of yielding probabilistic outcomes, which can help reduce uncertainty. Nevertheless, it has previously been observed that the effectiveness of the conventional Bayesian method is vulnerable to outliers. Therefore, a robust Bayesian technique has been devised to tackle this issue, and its effectiveness will be evaluated in comparison to the traditional Bayesian approach.

Kalman filter

The KF is a type of Bayesian filter that is used to estimate the recursive state of a dynamic system. This is achieved by reducing the MSE, even when there is process and

measurement noise present. The KF operates according to the following principles

$$\begin{aligned} \mathbf{xs}_t &= \mathbf{A}_t \mathbf{xs}_{t-1} + \mathbf{q}_t, \\ \mathbf{y}_t &= \mathbf{H}_t \mathbf{xs}_t + \mathbf{m}_t, \end{aligned} \tag{5.12}$$

where $\mathbf{xs}_t$ is the actual state at time t and $\mathbf{y}_t$ is its observation, the matrix $\mathbf{A}_t$ describes the state-transition model, $\mathbf{q}_t$ is the process noise, $\mathbf{H}_t$ is the observation matrix, which maps the actual state space into the observed space, and $\mathbf{m}_t$ is the observation noise. The noise terms $\mathbf{q}_t$ and $\mathbf{m}_t$ are assumed to be normally distributed: $\mathbf{q}_t \sim \mathcal{N}(0, \mathbf{Q}_t)$, $\mathbf{m}_t \sim \mathcal{N}(0, \mathbf{M}_t)$. In the literature, several approaches can be employed to extract KF formulation: for instance, Bayesian rule, maximum posterior approaches (MAP), orthogonal principle, and weighted least square method (WLS). To describe the state estimation procedure given the observations $\mathbf{y}_t$, we need to introduce the notation of $\hat{\mathbf{xs}}_{t|t}$ and $\hat{\mathbf{xs}}_{t|t-1}$ being the posterior and the prior estimates of the state $\mathbf{xs}_t$ given the observations till the time t and $t-1$, respectively. The cost function in WLS, defined as a sum of two squared quadratic norms [141], which combines the measurement (first term) and the prior estimate (second term) in finding the optimal posterior estimate of the $\mathbf{xs}_t$, is given as follows

$$J = \frac{1}{2}\|\mathbf{y}_t - \mathbf{H}_t \mathbf{xs}_t\|^2_{\mathbf{M}_t^{-1}} + \frac{1}{2}\left\|\mathbf{xs}_t - \mathbf{A}_t \hat{\mathbf{xs}}_{t-1|t-1}\right\|^2_{\mathbf{P}^{-1}_{t|t-1}}, \tag{5.13}$$

where -1 in the superscript denotes matrix inversion, quadratic norm of a vector $\mathbf{u}$ defined with a weights matrix $\mathbf{W}$ is the quadratic form obtained with the formula (denoting matrix transposition with T in the superscript)

$$\|\mathbf{u}\|_{\mathbf{W}} = \sqrt{\mathbf{u}^T \mathbf{W} \mathbf{u}} \tag{5.14}$$

and $\mathbf{P}_{t|t-1}$ is the prior estimate covariance matrix

$$\mathbf{P}_{t|t-1} = E[(\mathbf{xs}_t - \hat{\mathbf{xs}}_{t|t-1})(\mathbf{xs}_t - \hat{\mathbf{xs}}_{t|t-1})^T]. \tag{5.15}$$

The posterior estimate covariance matrix $\mathbf{P}_{t|t}$ is expressed in an analogous way as above

$$\mathbf{P}_{t|t} = E[(\mathbf{xs}_t - \hat{\mathbf{xs}}_{t|t})(\mathbf{xs}_t - \hat{\mathbf{xs}}_{t|t})^T]. \tag{5.16}$$

The KF formulation can be derived by minimizing J from Eq. (5.13) with respect to $\mathbf{xs}_t$, namely the posterior estimator of the state of the model is defined as

$$\hat{\mathbf{xs}}_{t|t} = \underset{\mathbf{xs}_t}{\operatorname{argmin}}\{J\}. \tag{5.17}$$

At the same time the covariance matrix is also estimated. The resulting recursive procedure is described with the following equations which are usually divided into two steps, namely the prediction step

$$\begin{aligned} \hat{\mathbf{xs}}_{t|t-1} &= \mathbf{A}_t \hat{xs}_{t-1|t-1}, \\ \hat{\mathbf{P}}_{t|t-1} &= \mathbf{A}_t \hat{\mathbf{P}}_{t-1|t-1} \mathbf{A}_t{}^T + \mathbf{Q}_t, \end{aligned} \tag{5.18}$$

and the update step

$$\begin{aligned} \hat{\mathbf{xs}}_{t|t} &= \hat{\mathbf{xs}}_{t|t-1} + \overline{\mathbf{K}}_t(\mathbf{y}_t - \mathbf{H}_t \hat{\mathbf{xs}}_{t|t-1}), \\ \hat{\mathbf{P}}_{t|t} &= (\mathbf{I} - \overline{\overline{\mathbf{K}}}_t \mathbf{H}_t)\hat{\mathbf{P}}_{t|t-1}(\mathbf{I} - \overline{\mathbf{K}}_t \mathbf{H}_t)^T + \overline{\mathbf{K}}_t \mathbf{M}_t \overline{\mathbf{K}}_t{}^T. \end{aligned} \tag{5.19}$$

The term $\overline{\mathbf{K}}_t$ in the above equations is called Kalman gain and for the standard KF its optimal value is given by the formula

$$\overline{\mathbf{K}}_t = \hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T(\mathbf{H}_t\hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T + \mathbf{M}_t)^{-1}. \tag{5.20}$$

As mentioned before, the KF gives the optimal solution when both the process noise and the measurement noise have a Gaussian distribution. For other distributions, the KF produces a sub-optimal solution, which is because only the second-order measurement information is used.

In the following subsection, the maximum correntropy Kalman filter (MCKF) is derived to handle non-Gaussian noise based on the following references [142, 141, 143, 144].

Maximum correntropy Kalman filter

In the following, for the model expressed with Eq. (6.5) another form of an objective function is introduced [141]. Based on the maximum correntropy criterion it has the form

$$J_{MC} = G_\sigma(\|\mathbf{y}_t - \mathbf{H}_t\mathbf{xs}_t\|_{\mathbf{M}_t^{-1}}) + G_\sigma(\|\mathbf{xs}_t - \mathbf{A}_t\hat{\mathbf{xs}}_{t-1|t-1}\|_{\mathbf{P}_{t|t-1}^{-1}}), \tag{5.21}$$

where $G_\sigma(u) = \exp(-\frac{u^2}{2\sigma^2})$ and σ is kernel size.

This function used instead of a previous cost function makes the estimation more robust in the presence of non-Gaussian noise, which is the case when the real observation process does not necessarily fulfill the assumptions of the KF model. The MCKF formulation can be derived by finding $\mathbf{xs}_t$ that maximizes the objective function J_{MC} given by Eq. (5.21). It should be emphasized that in the derivation based on WLS to obtain the solution given by a closed formula, an approximation is made, namely the posterior estimate is replaced at one place with the prior estimate, i.e. $\hat{\mathbf{xs}}_{t|t} \approx \hat{\mathbf{xs}}_{t|t-1}$ (please see [141] for details). The prediction and update steps are driven by the same equations as for the standard KF; see Eq. (5.18) and (5.19) but the optimal Kalman gain for the MCKF is given with the formula

$$\overline{\mathbf{K}}_t = \hat{\mathbf{P}}_{t|t-1}\lambda_t\mathbf{H}_t^T(\mathbf{H}_t\hat{\mathbf{P}}_{t|t-1}\lambda_t\mathbf{H}_t^T + \mathbf{M}_t)^{-1}, \tag{5.22}$$

where

$$\lambda_t = G_\sigma\Big(\|\mathbf{y}_t - \mathbf{H}_t\hat{\mathbf{xs}}_{t|t-1}\|_{\mathbf{M}_t^{-1}}\Big). \tag{5.23}$$

In the state estimation procedure using MCKF, when a significant outlier appears, the term $\mathbf{y}_t - \mathbf{H}_t\hat{\mathbf{xs}}_{t|t-1}$ in Eq. (5.19) (usually called innovation or pre-fit residual) diverges, but $\overline{\mathbf{K}}_t$ controls the divergence of the estimator $\hat{\mathbf{xs}}_{t|t}$. Please see references [142, 141, 143, 144] for more details about the procedure of driving equations and stability.

To illustrate the performance of MCKF, a two-state dynamic system is considered as follows

$$\mathbf{xs}_t = \mathbf{A}_t\mathbf{xs}_{t-1} + \mathbf{q}_t, \qquad \mathbf{y}_t = \mathbf{H}_t\mathbf{xs}_t + \mathbf{m}_t + t_t(\nu), \qquad \mathbf{A}_t = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad \mathbf{H}_t = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{5.24}$$

where Δt is the discrimination step size and, for this case, $\Delta t = 0.1$. The initial parameters of estimator are $\hat{\mathbf{xs}}_{0|0} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$, $\hat{\mathbf{P}}_{0|0} = \begin{bmatrix} 6.2\text{e}-08 & 1.2\text{e}-06; \\ 1.2-06 & 2.5-05 \end{bmatrix}$. Note

that in this special case y_t and m_t both reduce from a vector to a scalar. The noise terms $\mathbf{q}_t$ and $\mathbf{m}_t$ are normally distributed.
Where $\mathbf{q}_t \sim \mathcal{N}([0;0], [\frac{1}{4}(\Delta t)^4, \frac{1}{2}(\Delta t)^3; \frac{1}{2}(\Delta t)^3, (\Delta t)^2])$, $\mathbf{m}_t \sim \mathcal{N}(0, 10)$. The additional term $t_t(\nu)$ is a non-Gaussian noise generated as a series of of independent random variables from Student's t-distribution with number of degrees of freedom $\nu = 2.8$ (please note that this term is not included in the standard KF model).

The simulation results are plotted in Fig. 5.13. Panel (a) presents an actual state by a red line, and the observation of the real state includes heavy-tailed noise. As can be seen in panel (b) of Fig. 5.13, in the presence of impulsive noise, MCKF could properly track the mean state, while standard KF estimation was affected by heavy-tailed distributed noise and therefore could not follow the mean state with comparable accuracy.

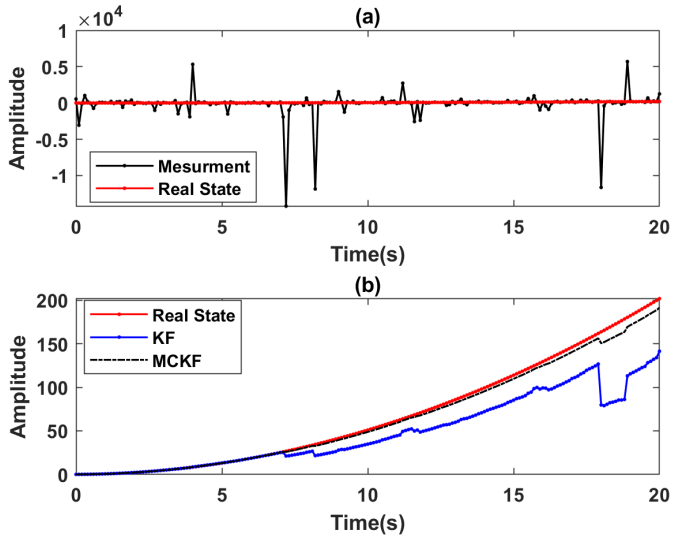


FIGURE 5.13: State estimation in the presence of non-Gaussian noise, (a) measurement signal and real state, (b) results of state estimation with KF and MCKF.

Switching maximum correntropy Kalman filter

In this subsection, the switching maximum correntropy Kalman filter (SMCKF) is introduced, which is a robust version of SKF. The SMCKF can be described as a dynamic Bayesian network in which, during the lifetime of the system, different KF or MCKF models are switched between. See Fig. 5.14. The additional process SM_t called switching variable is introduced. Its state space is finite, that is, it consists of k values: $1, 2, \ldots, k$. If $SM_t = j$, the degradation process is in the j-th of k subsequent stages and the process $\mathbf{xs}_t$ is in this stage modeled with the j-th of k MCKF models. In each time step, both the switching model variable SM_t and the state variable $\mathbf{xs}_t$ are hidden and thus must be figured out from observation $\mathbf{y}_t$. This may cause numerical problems, especially when the number of stages is increased, as argued in [145]. Kevin et al.in [146], developed an approximation approach, namely the generalized pseudo-Bayesian (GPB) method, to solve this issue.

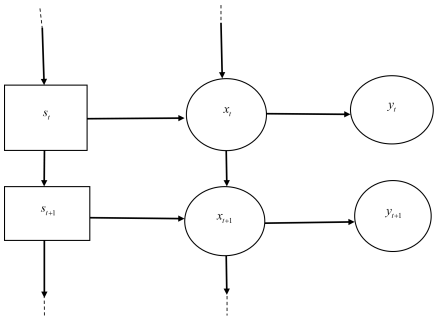


FIGURE 5.14: Dynamic Bayesian network.

Our goal in the segmentation through degradation stages is to evaluate, based on the measurement $\mathbf{y}_t = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$, the probabilities of the model being in a specific stage at each time t which directly corresponds to the values of

$$\pi_t^j = P(SM_t = j | \mathbf{y}_{1:N}) \tag{5.25}$$

for $j = 1, \ldots, k$. Note that the above probabilities are those filtered with the SMCKF based on measurement $\mathbf{y}_t$. For modeling the process SM_t, a discrete-time Markov chain is used. Thus, the non-filtered probabilities can be expressed with a recursive formula. Namely, the transition probabilities of SM_t, denoted as $z_{ij} = P(SM_t = j \mid SM_{t-1} = i)$ (assumed to be time-invariant), are given as

$$P(SM_t = j) = \sum_{i=1}^{k} P(SM_{t-1} = i) \cdot z_{ij}. \tag{5.26}$$

The framework used is similar to the one in [141] and it is based on the assumption that, in general, all parameters of the model, i.e., $\mathbf{A}_t$, $\mathbf{Q}_t$, $\mathbf{H}_t$ and $\mathbf{M}_t$, as well as the form of the state vector $\mathbf{xs}_t$, may explicitly depend on the current value of SM_t. In contrast to [141], it is assumed that in each of the 1 to k-th stages, the current HI needs to be estimated using MCKF instead of KF. Thus, the Kalman gain is calculated using Eq. (5.22), which contains the coefficient λ_t. Likewise, the covariance matrix of the innovation term $\boldsymbol{\epsilon}_t = \mathbf{y}_t - \mathbf{H}_t \hat{\mathbf{xs}}_{t|t-1}$ (given the measurement until time t) contains the same coefficient λ_t. Namely, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C_t})$ is obtained, where $\mathbf{C_t}$ is given by the formula

$$\mathbf{C}_t = \mathbf{H}_t \hat{\mathbf{P}}_{t|t-1} \lambda_t \mathbf{H}_t^T + \mathbf{M}_t. \tag{5.27}$$

As noted earlier, the parameters of the degradation model depend on SM_t describing the current stage. The estimates $\hat{\mathbf{xs}}_{t|t}$ and $\hat{\mathbf{P}}_{t|t}$, obtained by repeating iterations of SMCKF, so would be different for each possible series of $\{SM_1, \ldots, SM_N\}$. To reduce the computational complexity of the algorithm (and avoid iterating MCKFs through all possible trajectories of SM_t), the procedure is limited to iterating only through the possible values of the switching variable in the current and previous time, i.e., y_t and s_{t-1}. Therefore, the following form of the conditional probability density function is assumed

$$P(\boldsymbol{\epsilon}_t = u | SM_{t-1} = i, SM_t = j, \mathbf{y}_{1:t}) = g_{\mathbf{0},\mathbf{C_t}}(u), \tag{5.28}$$

where $g_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(u)$ is multivariate Gaussian probability density function of the random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Using Bayes' theorem conditioned on $\boldsymbol{\epsilon}_t$, the following is obtained

$$P(SM_{t-1}=i, SM_t=j|\boldsymbol{\epsilon}_t, \mathbf{y}_{1:t-1}) = \frac{g_{\mathbf{0},\mathbf{C_t}}(\boldsymbol{\epsilon}_t)\cdot P(SM_{t-1}=i, SM_t=j|\mathbf{y}_{1:t-1})}{\sum_{l=1}^{k}\sum_{m=1}^{k} g_{\mathbf{0},\mathbf{C_t}}(\boldsymbol{\epsilon}_t)\cdot P(SM_{t-1}=m, SM_t=l|\mathbf{y}_{1:t-1})}, \tag{5.29}$$

which holds for each pair of indices i and j. By summing these probabilities with respect to i (and using the definition of transition probabilities), a recursive formula for the filtered probability of the j-th MCKF model is obtained

$$\begin{aligned}\pi_t^j &= P(SM_t=j|\boldsymbol{\epsilon}_t, \mathbf{y}_{1:t-1})\\ &= \sum_{i=1}^{k} P(SM_{t-1}=i, SM_t=j|\boldsymbol{\epsilon}_t, \mathbf{y}_{1:t-1})\\ &= \frac{\sum_{i=1}^{k} g_{\mathbf{0},\mathbf{C_t}}(\boldsymbol{\epsilon}_t)\cdot \pi_{t-1}^i \cdot z_{ij}}{\sum_{l=1}^{k}\sum_{m=1}^{k} g_{\mathbf{0},\mathbf{C_t}}(\boldsymbol{\epsilon}_t)\cdot \pi_{t-1}^l \cdot z_{lm}}.\end{aligned} \tag{5.30}$$

This formula can be applied to calculate π_t^j at each step of the SMCKF.

To make the estimation procedure feasible at each time iteration, after the prediction and update steps (described in the previous subsections), an approximation step is added. This step merges the estimates, removing the dependence on previous values of the switching variable. Thus, only the estimates $\hat{\mathbf{xs}}_{t|t}$ and $\hat{\mathbf{P}}_{t|t}$ for $k\times k$ pairs of possible values of SM_{t-1} and SM_t need to be stored in memory. For the estimates calculated given the specific realization of the switching variable process, let us introduce the following notation: $\hat{\mathbf{xs}}_{t|t}|_{SM_{t-1}=i,SM_t=j}$ stands for the posterior estimate of $\mathbf{xs}_t$ obtained for $SM_{t-1}=i$ and $SM_t=j$.

The approximate value of the posterior estimate $\hat{\mathbf{xs}}_{t|t}$ given SM_t is evaluated taking a weighted mean of the state estimates obtained with each of the MCKS-s, i.e. starting with each of k possible values of k of SM_{t-1}. A similar formula is applied to the covariance matrices of the estimates. The two approximation formulas are the following

$$\tilde{\mathbf{xs}}_{t|t}|_{SM_t=j} = \sum_{i=1}^{k} w_t^{ij}\hat{\mathbf{xs}}_{t|t}|_{SM_{t-1}=i,SM_t=j}, \tag{5.31}$$

$$\tilde{\mathbf{P}}_{t|t}|_{s_t=j} = \sum_{i=1}^{k} w_t^{ij}(\hat{\mathbf{P}}_{t|t}|_{SM_{t-1}=i,SM_t=j} + \mathbf{r}\mathbf{r}^T), \tag{5.32}$$

where $\mathbf{r} = \tilde{\mathbf{xs}}_{t|t}|_{SM_t=j} - \hat{\mathbf{xs}}_{t|t}|_{SM_{t-1}=i,SM_t=j}$. Using Bayes' theorem and applying the results of Eq. (5.29) and (5.30), the proper weights are obtained

$$\begin{aligned}w_t^{ij} &= P(SM_{t-1}=i|SM_t=j, \boldsymbol{\epsilon}_t, \mathbf{y}_{1:t-1})\\ &= \frac{P(SM_{t-1}=i, SM_t=j|\boldsymbol{\epsilon}_t, \mathbf{y}_{1:t-1})}{P(SM_t=j|\boldsymbol{\epsilon}_t, \mathbf{y}_{1:t-1})}\\ &= \frac{g_{\mathbf{0},\mathbf{C_t}}(\boldsymbol{\epsilon}_t)\cdot \pi_{t-1}^i \cdot z_{ij}}{\sum_{l=1}^{k} g_{\mathbf{0},\mathbf{C_t}}(\boldsymbol{\epsilon}_t)\cdot \pi_{t-1}^l \cdot z_{lj}}.\end{aligned} \tag{5.33}$$

In this way, k estimates, after multiplying to $k\times k$ posterior estimates in prediction-update steps, reduce back to k values in the approximation step. The approximated values of the state estimate and covariance matrix (which overwrite the posterior) are then used as an input to the prediction step (obtaining new priors) in the next time iteration of the SMCKF.

More theoretical details on the procedure of the switching structure (in the case of SKF which is analogous to the and stability of the model) are available in [146]. The SMCKF, which is composed of multiple MCKFs with different state-space models, is applied to track changes in the degradation process. Then SMCKF is switched between the selected models according to their likelihood (Gaussian distribution of the innovation term) calculated from HI. Thanks to the use of the maximum correntropy criterion, the procedure is more robust than in the case of SKF. It should be noted that, for applying this approach, the degradation model should be selected based on the case's physics.

Based on the primary assumption, the degradation process comprises three stages. Three MCKF models are used to simulate the healthy, degradation and, critical stage. According to [147, 148] the polynomial MCKFs of zero, first, and second order are illustrated below, respectively. Using superscripts to denote the current value of the switching variable SM_t, the proposed models corresponding to each of the HI stages are presented. The model matrices and covariance matrices of the model noise are as follows

$$\mathbf{A}_t^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_t^2 = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_t^3 = \begin{bmatrix} 1 & \Delta t & \frac{(\Delta t)^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.34}$$

$$\mathbf{Q}_t^1 = \beta \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{Q}_t^2 = \beta \begin{bmatrix} \frac{(\Delta t)^3}{3} & \frac{(\Delta t)^2}{2} & 0 \\ \frac{(\Delta t)^2}{2} & \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{Q}_t^3 = \beta \begin{bmatrix} \frac{(\Delta t)^5}{20} & \frac{(\Delta t)^4}{8} & \frac{(\Delta t)^3}{6} \\ \frac{(\Delta t)^4}{8} & \frac{(\Delta t)^3}{3} & \frac{(\Delta t)^2}{2} \\ \frac{(\Delta t)^3}{6} & \frac{(\Delta t)^2}{2} & \Delta t \end{bmatrix}, \tag{5.35}$$

where Δt is the discrimination step of the process and observations (constant for data) and β is a scalar hyper-parameter (related to the noise) that can describe the uncertainty of the filter in actual application, which can be used for tuning of the SMCKF for a different machine.

Observation matrices corresponding to polynomial MCKFs of zero, first, and second order are defined as follows

$$\mathbf{H}_t^1 = \mathbf{H}_t^2 = \mathbf{H}_t^3 = \begin{bmatrix} 1, & 0, & 0 \end{bmatrix}. \tag{5.36}$$

The measurement noise term is a scalar in proposed application, and its variance (reducing $\mathbf{M}_t$ to a scalar) is fixed and independent on SM_t. For each case, different values are used depending on the particular scenario (simulation, real data analysis), as described in the following sections. The matrix containing the transition probabilities z_{ij} is fixed and defined as follows [91, 148]

$$\mathbf{Z} = \begin{bmatrix} 0.998 & 0.001 & 0.001 \\ \sim 0 & 0.999 & 0.001 \\ \sim 0 & \sim 0 & 1 \end{bmatrix}. \tag{5.37}$$

Note that, as the subdiagonal elements of $\mathbf{Z}$ are approximately equal to 0, there is no possibility of transition to the previous state (zero value is not explicitly assigned only to avoid numerical problems). The rate of degradation is supposed to only progress from healthy to degradation and critical stages, as it is common characteristic for a degradation process that the machine does not spontaneously repair. The superdiagonal elements of $\mathbf{Z}$ were arbitrarily chosen (all set as 000.1) to somehow

render the rare but inevitable event of the transition to the regime of larger degradation. For each regime, the biggest probability is naturally assigned to diagonal elements which are related to the event of remaining in the current regime.

The vector containing initial probabilities of each stage or MCKF (i.e π_0^1, π_0^2 and π_0^3) is set as [91, 148]

$$\boldsymbol{\pi}_0 = \left[\ 0.9\,,\quad 0.05\,,\quad 0.05\ \right]. \tag{5.38}$$

The first component of $\boldsymbol{\pi}_0$ has the biggest value, which stands for another natural assumption: at the beginning the machine was most likely healthy.

The initial estimate state of the model and its covariance matrix have the following form, respectively

$$\tilde{\mathbf{xs}}_{0|0} = \begin{bmatrix} y_0 \\ 0 \\ 0 \end{bmatrix}, \qquad \tilde{\mathbf{P}}_{0|0} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.39}$$

where y_0 is the initial measurement (a scalar in this case).

5.2.2 Simulation

In this part, initially, the long-term data is generated using the assumptions outlined in Chapter 3. Next, the suggested approach is utilized to divide the data into three distinct stages for two separate scenarios: 1) when there is Gaussian noise and 2) when there is Student's t-distributed noise , see Appendix A.

Result for Gaussian distributed model

The proposed methodology is implemented in this subsection using the data generated by the recommended model. It is assumed that the noise term follows a Gaussian distribution, with $\hat{R}(t)$ being distributed as $\mathcal{N}(0,1)$. The segmentation results for the simulated data are collectively shown for all of the chosen methods. At Time=1000, the healthy stage transitions to the deterioration stage (referred to as changing point 1 or CP1). The border between the degradation stage and the critical stage (referred to as the changing point 2 or CP2) occurs at Time=1600.

The findings of the suggested approach, together with the standard SKF, are provided in Fig. 5.15. Panel (a) displays the HI and the actual point of change. Panels (b) and (c) show the likelihood of each step (i.e. $\pi_t^1, \pi_t^2, \pi_t^3$) executed by SKF and SMCKF, respectively. The colors green, yellow, and red correspond to the stages of health, degeneration, and criticality, respectively. On these panels, it is evident that from time = 0 to 1000, the likelihood of the model being in the healthy stage is the largest compared to the other two stages. However, as time approaches 1000, the probability of the healthy stage decreases while the probability of the degradation stage grows. From Time=1000 to Time=1600, the degradation stage consistently has the highest likelihood stage. After the time reaches 1600, the critical stage has the greatest likelihood. As anticipated, when Gaussian noise is present, both approaches yield nearly identical results. Furthermore, panels (d) and (e) show the state that is most likely to occur during the deterioration process, as well as the points at which the phases change.

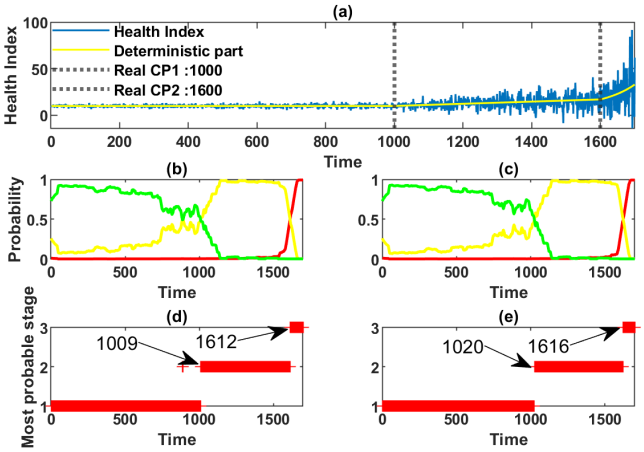


FIGURE 5.15: Detection of the stages in the presence of Gaussian noise, (a) HI, (b) probability of stages performed by SKF, (c) probability of stages performed by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF.

Result for non-Gaussian distributed model

In this section, the suggested approach is implemented on data generated by the provided model, assuming that the noise component follows the Student's t distribution with a specified degree of freedom $v = 3$. Based on the simulation technique, CP1 and CP2 have values of Time=1000 and Time=1600, respectively.

Fig. 5.16 displays the outcomes of the suggested techniques in non-Gaussian noise conditions. Panels (b) and (c) display the estimated probabilities of each condition using SKF and MCSKF, respectively, throughout the deterioration process. In panel (b), the probability of being in the healthy stage (represented by the green line) begins to decrease starting at Time=300. After some fluctuations with the degradation stage (represented by the yellow line), around Time=880, the degradation stage becomes the most probable stage and remains so until Time=1446. At this point in time, the critic stage has a higher likelihood than the other two stages. However, as depicted in panel (b) of Fig. 5.16, the SKF has deviated from the expected path after Time=1500. Panel (c) displays the outcome of the SMCKF. The panels (d) and (e) display the most likely stage in the degradation process and the points of change identified using SKF and SMCKF. Upon comparing the two final panels, it is evident that the SMCKF result is more proximate to the actual changing points, as anticipated. However, SKF has been greatly impacted by non-Gaussian noise and ultimately deviates, rendering the results from SKF unacceptable.

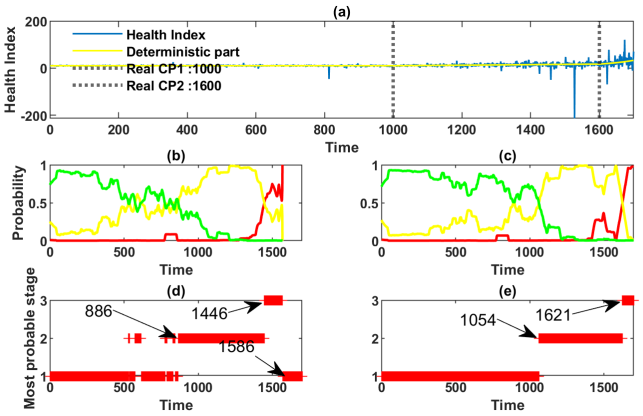


FIGURE 5.16: Detection of the stages in the presence of non-Gaussian noise, (a) HI, (b) probability of stages by SKF, (c) probability of stages by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF.

The procedure was replicated 100 times using the same model but with varying levels of non-Gaussian noise (represented by different degrees of freedom, ν). The sensitivity and specificity values were calculated for each simulation, and the mean values were calculated. These mean values are presented in Fig. 5.17 and Fig. 5.18 for sensitivity and specificity, respectively. Furthermore, it is important to mention that the kernel size value is consistently chosen as $\sigma = 0.391$ for all simulations.

The sensitivity results are shown in Fig. 5.17. Panel (a) clearly demonstrates that the SKF technique outperforms the suggested method during the healthy stage. However, the discrepancy is minimal, and both methods yield values close to one over different ranges of ν. The sensitivity results for the degradation stage are shown in panel (b). It is evident that the results for SMCKF remain rather consistent with different values of ν, with a sensitivity value greater than 0.9. In contrast, the sensitivity of SKF is significantly lower for lower values of ν. In addition, panel (b) shows that the sensitivity of the SKF is enhanced by increasing the value of ν, indicating the impact of the non-Gaussianity level on the performance of the SKF. Furthermore, the circumstances in the last stage (referred to as the crucial stage) are identical to those in the degradation stage, as depicted in panel (c). The sensitivity of the SMCKF is greater (approximately one) and more consistent than that of the SKF during the final stage. Moreover, it is important to note that based on the findings depicted in Fig. 5.17, specifically in the final two panels, the SMCKF exhibits a superior performance compared to the SKF, even when the noise distribution is completely similar to the Gaussian distribution (for $v = 100$).

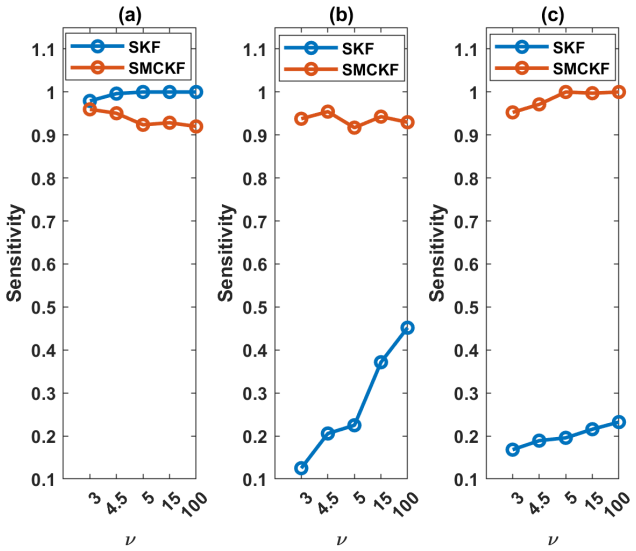


FIGURE 5.17: Sensitivity analysis, (a) health stage, (b) degradation stage, (c) critical stage.

The results of specificity are shown in Fig. 5.18. As observed in panels (a) and (b), the SMCKF exhibits a greater (roughly close to one) and more consistent specificity value compared to SKF. Furthermore, in each of the aforementioned panels, the specificity of the SKF is significantly diminished for smaller values of ν. Similar to sensitivity, the impact of the non-Gaussianity level on the performance of SKF is apparent. Furthermore, at the last step (referred to as the critical stage), the SKF exhibits superior performance compared to the SMCKF. However, the disparity between the two methods is minimal, and both approaches demonstrate a specificity near to 1 for all ν values.

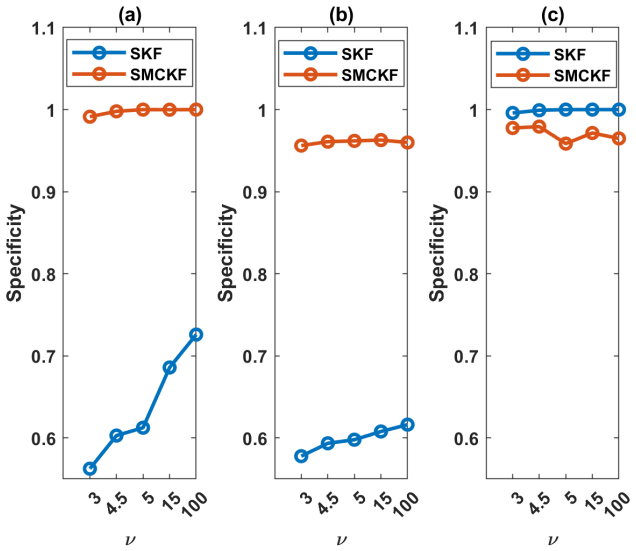


FIGURE 5.18: Specificity analysis, (a) health stage, (b) degradation stage, (c) critical stage.

5.2.3 Real data analysis

In this section, the proposed methodology is verified using existing real datasets.

Result for FEMTO dataset

Fig. 5.19 presents the results of the SMCKF and the SKF applied to the FEMTO dataset. The probability of the phases (i.e. $\pi_t^1, \pi_t^2, \pi_t^3$) during the deterioration process is depicted in panels (b) and (c) using SKF and SMCKF, respectively. The panels (d) and (e) depict the most likely stage detected and the points at which the stages change, as determined by SKF and SMCKF. Upon comparing the outcomes of SKF and SMCKF in the final two panels, it becomes evident that the results are largely similar. However, in panel (d) where SKF is depicted, there is noticeable fluctuation around Time=1300. Furthermore, in both approaches, the crucial phase is discovered at Time=2411, indicating that it is recognized prior to the actual point of transition.

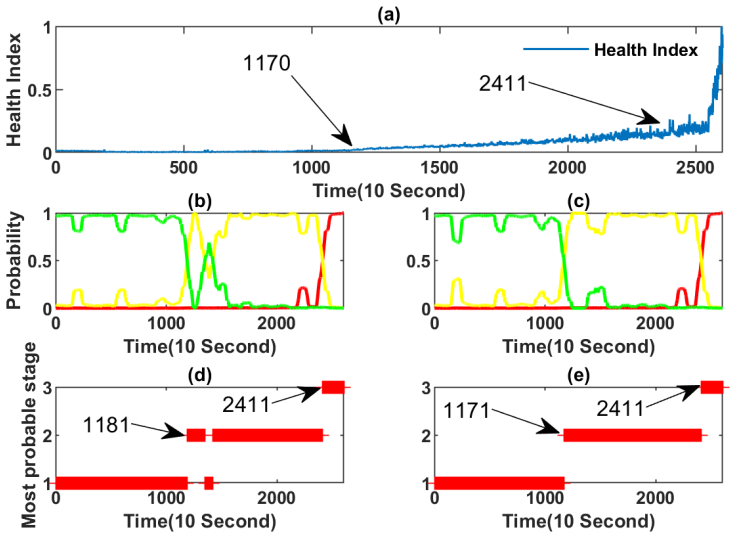


FIGURE 5.19: Segmentation of the FEMTO datasets, (a) HI, (b) probability of stage by SKF, (c) probability of stage by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF.

Results for Wind turbine dataset

The wind turbine dataset showcases the outcomes of the SMCKF and SKF, as depicted in Fig. 5.20. Panel (b) displays the probability of phases in the degradation process according to SKF, while panel (c) shows the probability based on SMCKF. Furthermore, panels (d) and (e) depict the most likely stage identified by SKF and SMCKF, as well as the places at which the stages transition. Upon comparing the results of SKF and SMCKF in the last two panels, it is evident that both approaches identified 18-Mar as the initial change point. However, SKF detected the second changing point on 9-Apr, whereas SMCKF saw this point on 21-Apr.

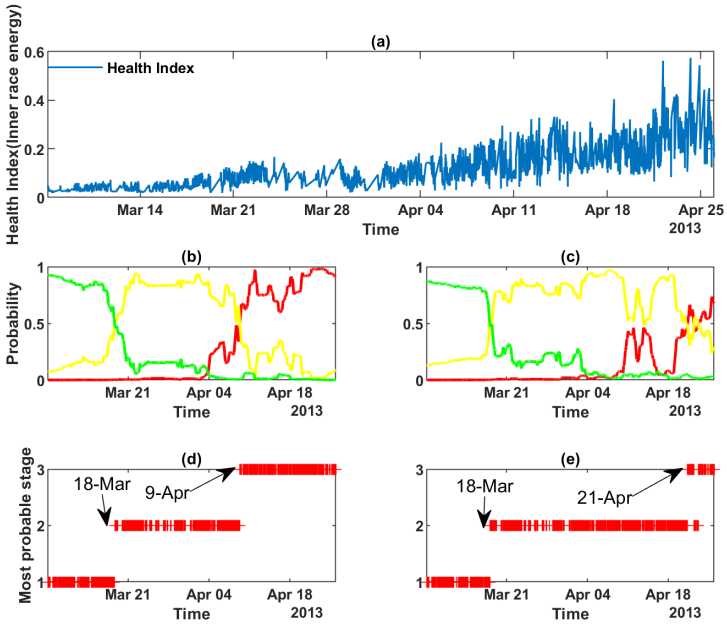


FIGURE 5.20: Segmentation of the wind turbine datasets, (a) HI, (b) probability of stage by SKF, (c) probability of stage by SMCKF, (d) most probable stages based on the implementation of SKF, (e) most probable stages based on the implementation of SMCKF.

As mentioned previously, this dataset consists of two versions. The second version consists a raw vibration signal that is recorded for six seconds every day, with a sampling rate of 100k. Please refer to panel (a) in Fig. 5.11 for more details. Therefore, it possesses a remarkable capacity to utilize diverse techniques of frequency analysis for the purpose of conducting condition monitoring. The envelope analysis is performed on each set of vibrations, and the results are then presented sequentially in panel (b) of Fig. 5.11. Panel (b) in Fig. 5.11 illustrates the presence of a harmonic frequency after 18 March, indicating that the bearing has entered the degradation stage. Furthermore, following the date of April 22nd, there is a significant increase in the amplitude of the harmonic frequency, which can be classified as CP2. Upon comparing the findings of the SMCKF and SKF methods with the envelope analysis, it is evident that CP1 and CP2, as identified by the SMCKF (CP1 = 18 March and CP2 = 22 April), correspond closely to the CP1 and CP2 recognized through visual inspection in the envelope analysis.

Moreover, in this chapter, the kernel size σ was determined through a trial and error process, which is a limitation of the suggested methodology. However, this issue is acknowledged and planned to be addressed in future research.

5.3 Chapter highlights

In this chapter, the segmentation of long-term data using both offline and online approaches is explored. In the first part of this chapter, the regression approach is employed to divide the HI into three different segments based on its deterministic behavior in the presence of varying levels of non-Gaussian noise. The aim is to demonstrate the effect of this type of noise on segmentation performance. The results indicate that robust methods, both in the simulated and real cases, yield superior outcomes.

While the first part of this chapter showcases the potential for segmenting the HI based on deterministic behavior, the second part focuses on developing a robust online Bayesian approach for segmenting the HI, making it more applicable for real-world scenarios.

Segmenting health indices into three regimes (health, slow degradation, and fast degradation) based on their dynamic behavior, especially in the presence of non-Gaussian (heavy-tailed) noise, offers several practical advantages in industries where data is affected by non-Gaussian noise, such as mining. These robust proposed approaches focus on dynamic trends rather than fixed thresholds. By doing so, the system can reduce the frequency of false alarms that commonly occur in traditional threshold-based methods. In addition, the development of a robust approach for segmenting the HI into different regimes in this chapter could pave the way for the creation of simpler prediction models in the subsequent chapters.

Chapter 6

Predicting of remaining useful life based on long-term health index

In this chapter, the prediction of RUL is explored, particularly in the presence of non-Gaussian noise. Previous chapters have laid the groundwork by demonstrating how to model, identify, and segment the HI, while also thoroughly investigating the effects of non-Gaussian noise. Therefore, this chapter shifts its focus to prediction, a crucial aspect for decision-making, PHM applications.

6.1 Probabilistic methodology for predicting remaining useful life

This section delineates the methodology founded on a degradation process for estimating the RUL. The RUL estimation workflow, as proposed approach, is depicted in Fig. 6.1. The acquired data, such as sensor-recorded vibration data, act as indicators for monitoring the machine's health condition. Using signal processing techniques, a HI is constructed from the measured signals, effectively representing the machine's health status. Based on the evolving degradation patterns of the HI, the life of the machinery is segmented into three distinct stages: healthy, degradation, and critical stages, with the FPT marking the initial entry into the critical stage.

During the critical stage, a state-space degradation model is introduced, with its parameters estimated using the MCEKF and updated in real time as new HI values become available. An EOL threshold is defined, facilitating the calculation of the PDF of the estimated RUL based on the constructed degradation model and the EOL threshold.

Furthermore, the details of the state-space degradation model and the theories behind it, including the derivation of the EKF and MCEKF, are explained in the following subsections.

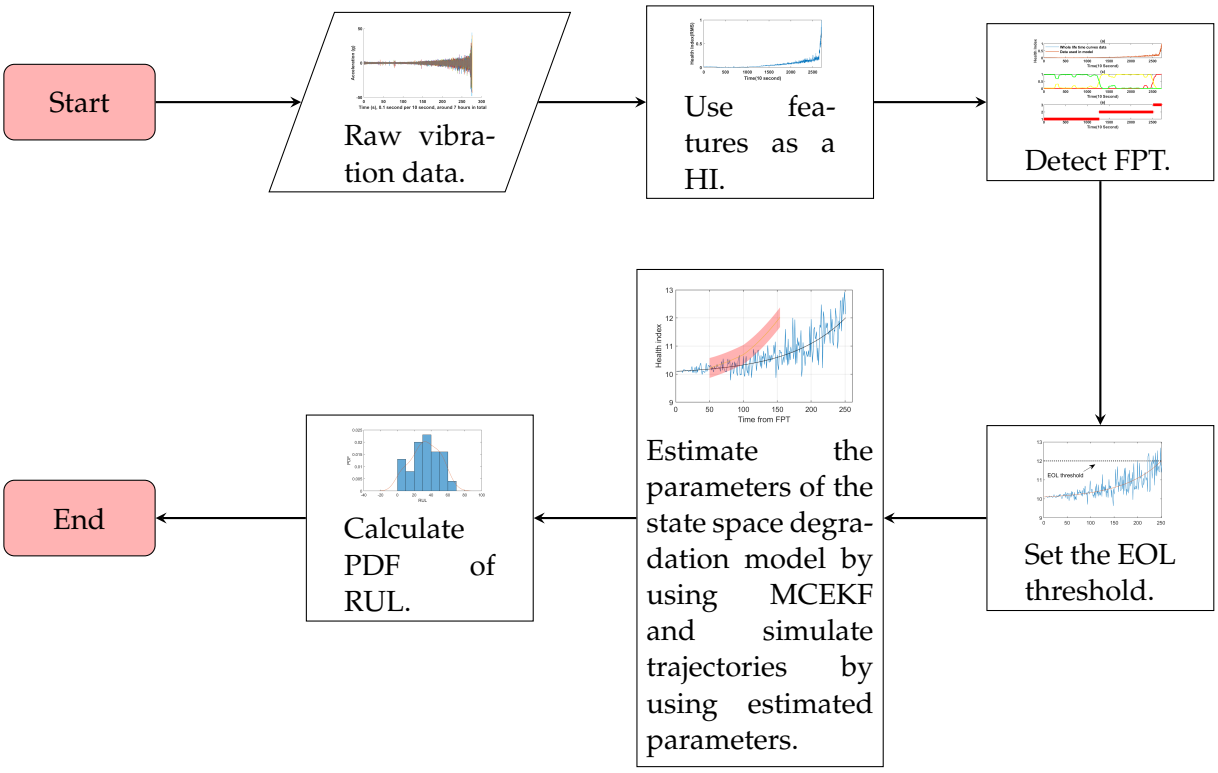


FIGURE 6.1: Complete procedure for estimation of the distribution of RUL.

6.1.1 Remaining useful life

The RUL is a key metric in machine prognostics, indicating the expected operational duration before repair or replacement is needed. Fig. 6.2 provides a visual representation of RUL. Our methodology emphasizes training a predictive model exclusively with historical data. Subsequently, this model is utilized to forecast future degradation process values, accompanied by CI. The estimated degradation curve is analyzed to identify the moment when it initially exceeds a predefined EOL threshold, denoted t_{EOL}. The predicted RUL is then calculated as the difference between this failure time t_{EOL} and the last observation time in the training dataset (t_0), represented as $t_{\text{EOL}} - t_0$.

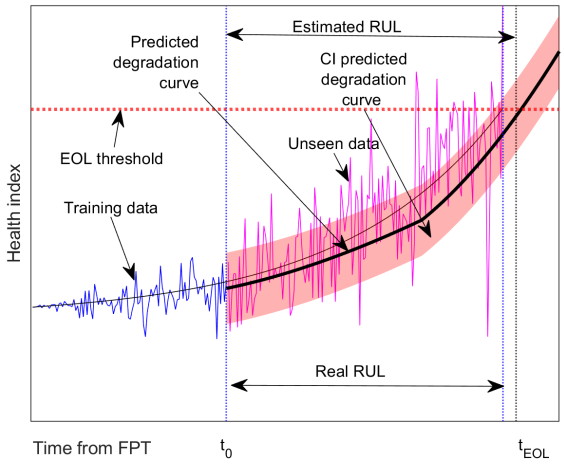


FIGURE 6.2: Concept of RUL of the equipment.

6.1.2 State space degradation model

As confirmed by the findings in [149], the exponential model consistently demonstrates a better global regression performance compared to polynomial and other conventional models. Consequently, the ensuing exponential model is adopted as the degradation pattern designated for the analysis of the third regime of the degradation process (critical stage)

$$S_t = a\exp(bt) + c, \tag{6.1}$$

where S_t corresponds to last stage of the HI model, see section 3.2, t is the discrete time and a, b, c represent the parameters of the model that need to be initialized with real data.

Recognizing the intricate and time sensitive nature of the degradation process, this study postulates that the model parameters a, b, c, are subject to temporal variations. This assumption is closely aligned with the dynamics of the degradation process in the real world. Let the dynamic state $\mathbf{xs}_t$ be defined as a column vector $\mathbf{xs}_t = [a, b, c]^T$ and y_t is observation. Furthermore, recognizing the impact of external noise and inherent uncertainty, the subsequent state-space model is introduced

$$\begin{cases} \mathbf{xs}_t = \mathbf{xs}_{t-1} + \mathbf{q}_t \\ y_t = a\exp(bt) + c + m_t \end{cases}, \tag{6.2}$$

where, the process noise is denoted as $\mathbf{q}_t$, presumed to be a vector white noise, i.e. series of independent identically distributed zero-mean random vectors. This noise factor is characterized by the covariance matrix $\mathbf{Q}_t$. Under no explicit assumption about the distribution, a general location-scale distribution is assumed for $\mathbf{q}_t$ at each t, denoted as: $\mathbf{q}_t \sim p(\mathbf{0}, \mathbf{Q}_t)$; the first parameter is the mean and the second parameter is the squared scale in this notation. In addition, m_t is assumed to be a scalar white noise (i.e., a series of independent identically distributed zero-mean scalar random variables), not necessarily Gaussian, to suit the case of heavy-tailed noise inherent in real data (in the Kalman filter framework it is called the measurement noise; however, in this work, it does not have such an explicit interpretation). Noise m_t is characterized by a parameter of the time scale M_t. Here, let us only denote

(analogously to $\mathbf{q}_t$) a general distribution of m_t, i.e. $\mathbf{m}_t \sim p(0, M_t)$, for each t. Keeping an analogy with the degradation model from Eq. (6.5), the relationship between HI and the parameters featured in Eq. (6.2) is mapped, denoting it with the non-linear functions $\mathbf{f}$ and h as follows

$$\mathbf{f}(\mathbf{xs}_{t-1}) = \mathbf{xs}_{t-1}, \tag{6.3}$$

$$h(\mathbf{xs}_t) = a_t \exp(b_t t) + c_t. \tag{6.4}$$

Furthermore, $\hat{\mathbf{xs}}_0$ – the initial state estimate, and $\hat{\mathbf{P}}_{0|0}$ – the initial estimate for the posterior covariance matrix, are set. Additionally, all values of $\mathbf{Q}_t$ and M_t need to be defined as well. In the following two subsections, EKF is described and MCEKF is introduced, considering more general notation. Please note that in the general approach, y_t would be replaced with a vector $\mathbf{y}_t$, the scalar function h with a vector function $\mathbf{h}$, and the scale parameter M_t would be replaced with the matrix $\mathbf{M}_t$, which refers to the covariance matrix in the Gaussian case.

6.1.3 Extended Kalman filter

The EKF works based on the use of Taylor series to transform non-linear filtering problems into linear forms. The discrete non-linear state-space representation of the model can be presented as follows

$$\begin{aligned} \mathbf{xs}_t &= \mathbf{f}(\mathbf{xs}_{t-1}) + \mathbf{q}_t, \\ \mathbf{y}_t &= \mathbf{h}(\mathbf{xs}_t) + \mathbf{m}_t, \end{aligned} \tag{6.5}$$

where $\mathbf{xs}_t$ is the unknown dynamic state at time t and $\mathbf{y}_t$ is its observation, respectively – both of them are in general vectors of selected size. Furthermore, $\mathbf{q}_t$ and $\mathbf{m}_t$ represent processes and measurement noise that are considered Gaussian distributed: $\mathbf{q}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$, $\mathbf{m}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{M}_t)$. The derivation process follows the same steps as outlined in Section 5.2.1, but the MCEKF employs the following Eq.(6.6) and Eq.(6.7) to linearize the system and the rest is the same

$$\mathbf{A}_t = \frac{d\mathbf{f}}{d\mathbf{xs}}\Big|_{\hat{\mathbf{xs}}_{t-1|t-1}}, \tag{6.6}$$

$$\mathbf{H}_t = \frac{d\mathbf{h}}{d\mathbf{xs}}\Big|_{\hat{\mathbf{xs}}_{t|t-1}}. \tag{6.7}$$

6.1.4 Discussion of the hyperparameters setting

In the state estimation procedure using MCEKF, the term $\mathbf{y}_t - \mathbf{h}(\hat{\mathbf{xs}}_{t|t-1})$ in Eq. (5.19) (also known as the innovation or prefit residual) diverges when a significant outlier occurs. However, the estimator $\hat{\mathbf{x}}_{t|t}$ is controlled by $\mathbf{K}_t$ to prevent divergence. For more information on the stability issue, refer to the following sources: [142, 141, 143, 144].

Furthermore, it is important to note that in the MCEKF, the σ parameter, also known as kernel size, determines the level of sensitivity of the filter to discrepancies between the predicted and measured states. A higher value of σ increases the filter's tolerance to deviations, whereas a lower value of σ enhances its sensitivity to minor variations. This increased sensitivity can help to reject outliers, but it may also make the filter more susceptible to noise. Hence, the careful selection of an appropriate value for σ is vital to achieve a balance between resistance to noise and sensitivity to

variations in the data. Determining the optimal value of σ for a certain application typically requires testing or expertise in the field.

As mentioned in the reference [150], the process error, denoted as $\mathbf{q}_t$, encompasses the uncertainty related to the filter's ability to represent real-world dynamics. A reduction in process error improves the filter's ability to accurately fit condition monitoring data, increasing the model's sensitivity to subtle changes. However, this may also make the model more prone to overfitting. Therefore, it is crucial to appropriately allocate the values for the members of the matrix $\mathbf{Q}_t$, which represent the noise parameters that affect the system dynamics. This matrix is designated as a constant covariance matrix, denoted as $\mathbf{Q}_t = \mathbf{Q}$. The diagonal nature of the matrix is a result of assuming uncorrelated (white) noise. The off-diagonal entries are zero, suggesting that there is no correlation between distinct state variables. The diagonal elements represent the variances of each individual state variable. It is important to note that the noise is not uniform in all directions - the diagonal elements might differ from each other. This is because the healthy, degraded, and critical stages have distinct characteristics. Choosing modest values for the individual diagonal members of the covariance matrix indicates a low level of uncertainty in the dynamics of the process. This suggests that a good understanding of the system's behavior or minimal noise affecting the system is initially assumed. These numbers were determined by analyzing historical data obtained from comparable cases. For detailed insights into the ability to analyze past degradation data, refer to earlier publication [149]. However, considering the fluctuations in material qualities on the microscopic scale, working conditions, and environmental influences, it may be necessary to modify the matrix $\mathbf{Q}$ individually, potentially by enhancing a particular diagonal member.

Similarly, the measurement error, represented as m_t, is found by computing the variance of the stationary measurements obtained while the machine is in normal working condition. The specific constant value of the measurement error scale parameter, denoted as $M_t = M$, may vary between different scenarios. The validity of assuming stationary measurements from well-functioning machinery, as utilized in this study, is thoroughly confirmed by the analysis of data from many examples, hence affirming its practical application.

The initial covariance matrix, denoted $\mathbf{P}_{0|0}$, is arbitrarily chosen with an identity matrix. The MCEKF model for the prognostic evaluation of bearing degradation is constructed by establishing matrices and defining initial values.

6.1.5 End of the life of machine's equipment

The EOL in machinery is a critical parameter and selecting the right threshold is paramount for effective maintenance and operation. It marks the point at which a machine is expected to become unreliable, which makes it crucial for scheduling preventive maintenance, cost reduction, safety, resource optimization, reliability, and environmental impact. The key to success lies in data-driven approaches, such as predictive maintenance techniques and thorough data analysis, which enable organizations to accurately identify and set appropriate EOL thresholds for their machinery. It should be noted that choosing an appropriate EOL value for real datasets presents a considerable challenge, and this has been the subject of extensive research over the past decade [56, 151]. The complexity of this task endures.

The main objective of this study is not to choose a particular EOL criterion for the degradation process. Alternatively, the proposed method can be used to compute the RUL for various EOL levels. For simulated datasets, the EOL threshold is established as the magnitude of the endpoint of the deterministic component, and the

RUL is determined using this threshold. In the actual case study, where it is difficult to choose the threshold, it was selected using historical data. Specifically, the amplitude of one of the points in the last stage was determined as the EOL threshold, and the RUL was then estimated correspondingly. However, it should be acknowledged that the choice of threshold could potentially be determined using a different method, and the specific selection of the EOL threshold would not influence the overall context of this research work. The aim of this study is to present a robust method that minimizes the impact of outliers, such as heavy-tailed noise, on the prediction of HI using a Bayesian approach.

6.2 Simulation and results

6.2.1 Results for simulated signal

In this subsection, proposed methodology is applied to simulated data based on Subsection 6.2.1. the HI data is generated from the last stage (exponential part). The simulation is restricted only to the third regime starting from the FPT. To demonstrate the performance of the proposed method and to compare it with other filters from the KF family, the classic EKF and UKF were applied as well-known MCKF for non-linear state estimation. For more details on EKF and UKF, the reader is encouraged to refer to [152, 153].

Degradation process in the presence of the Gaussian noise

The degradation curves is simulated based on Subsection 6.2.1 from last regime. Also, the distribution of noise is consider Gaussian. The final results of applying the proposed method to the simulated data are presented in Fig. 6.3. The top panel illustrates the degradation curves and the EOL level. In each step, the data sequence originating from the FPT onward to the specific time point t was used to iteratively refine the model and derive accurate predictive estimates of the RUL. The bottom panel shows the estimated mean of RUL using unscented kalman filter (UKF), see [152] for more details, EKF, and the proposed MCEKF. At the same time the difference in time between the EOL point and the specific data point was calculated to represent the real RUL at that point. The black dashed line and the purple area show the real RUL value and the accuracy bound $\pm 20\%$, respectively, calculated for each data point t from FPT onward.

As expected, when the dataset size is small, the accuracy of RUL estimation using all filters is sub-optimal. As depicted in the lower panel, both the UKF and EKF tend to underestimate RUL, while the MCEKF tends to overestimate it. This discrepancy highlights the limitation of data-driven estimation methods when data availability is limited. After a while, around $t = 180$, it can be seen that MCEKF is approaching the accuracy bound and approximately remains in the accuracy bound until the end of the life. Meanwhile, the performance of UKF and EKF has improved with time. The results of these two methods are close to, however, until $t = 430$ not entirely within the accuracy bound. After that, their performance is approximately same as the results that performed by MCEKF approach and it is acceptable. Furthermore, it should be mentioned that the kernel size for MCEKF was determined as $\sigma = 0.5$ through a trial and error process, which serves as a crucial hyperparameter for the MCEKF algorithm. The selected σ value lies within an arbitrary range above zero, and it is noteworthy that when a large value for the σ is selected, the trends of both

the MCEKF and the EKF align. It is important to highlight that selecting a lower value of σ enhances the robustness of the model against outlying data points.

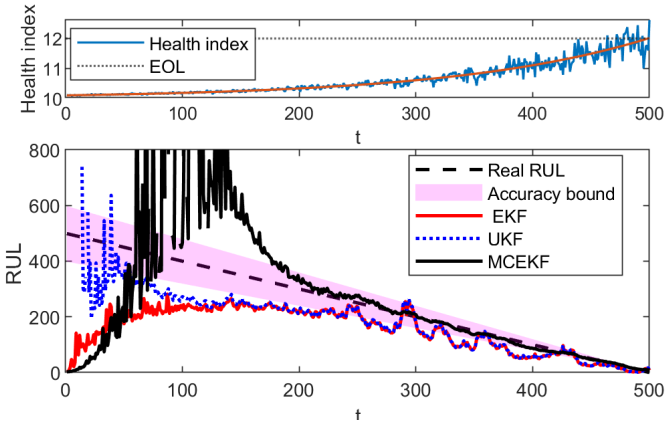


FIGURE 6.3: Final results of predicted RUL in the presence of the Gaussian noise, top panel: HI and EOL, bottom panel: predicting RUL results performed by EKF, UKF, and proposed MCEKF, and the real RUL ±20% accuracy bound.

The PDF changing over time, portraying the estimated RUL values obtained by the means of the MCEKF technique, is shown in Fig. 6.4. Evidently, in the initial stages, where the dataset contains an insufficient number of data points, the PDF of the estimated RUL does not yield favorable outcomes. However, a discernible improvement becomes apparent as the size of the dataset increases, allowing the PDF to progressively enclose the real RUL. In particular, beyond the temporal threshold of $t = 200$, the mean of the PDF aligns closely with the actual values of the RUL. Moreover, the 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 6.5. As can be seen; after $t = 200$, the RUL values are within the scope of 90% CI. This notable alignment between the real RUL and the 90% CI serves as evidence for the efficacy of proposed method, especially in the context of its robustness against the influence of the Gaussian noise.

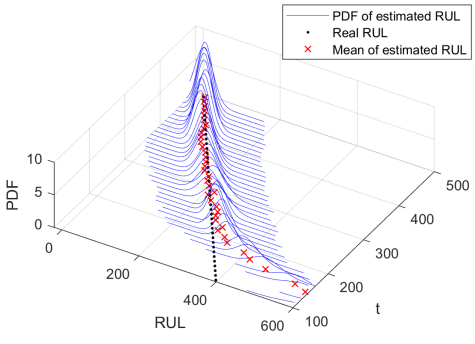


FIGURE 6.4: PDF of the estimated RUL in the presence of the Gaussian noise, changing over time.

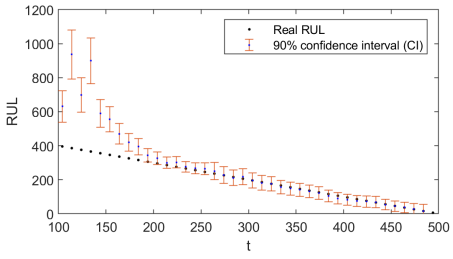


FIGURE 6.5: The 90% CI of RUL estimated using MCEKF for the simulated HI in the presence of Gaussian noise.

Degradation process in the presence of the non-Gaussian noise

The simulated degradation curve for the last regime based on Chapter 3 in the presence of non-Gaussian noise (Student's t with $\nu = 2.1$) and a time-varying exponential scale is shown in Fig. 6.6. Panel (a) presents HI and the deterministic trend of the HI. The random component is demonstrated in panel (b). As can be seen, the scale of the random component is exponentially growing over time. In panels (c) and (d), the scale and generated non-Gaussian noise used for the simulated degradation curve are shown.

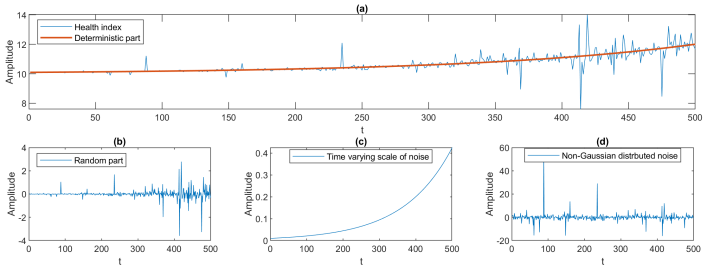


FIGURE 6.6: Simulated HI in the presence of non-Gaussian noise.

The final results of applying the proposed method to the simulated data are presented in Fig. 6.7. The top panel shows the degradation curves and the level of EOL. In the top panel, there are some strong outliers (related to non-Gaussian noise) around $t = 240, 380, 430, 480$ are highlighted on the degradation curve. One can expect that they may affect the non-robust methods. The bottom panel shows the estimated RUL using EKF, UKF, and the proposed MCEKF. The black dashed line and the purple area show the real RUL value $\pm 20\%$ accuracy bound (assigned for the assessment of the RUL prediction). As previously, when the number of data is lower, the precision of the estimated RUL for all filters tested is low, according to the bottom panel, RUL was underestimated by EKF and UKF, and overestimated by MCEKF. It can be seen that after a while, approximately starting from $t = 150$, RUL estimated by MCEKF remains in the accuracy bound until the end of the life. At the same time, the performance of UKF and EKF has improved over time and their results are close to the accuracy bound. However, until $t = 180$ the obtained trajectories are beyond the accuracy bound and after that the performance improved. Nevertheless, they are inferior to the MCEKF's result. Also, it can be clearly seen that in the case of the noise having non-Gaussian characteristics (top panel), the performance of the UKF and EKF is dramatically reduced (see the circles in the top and bottom panel). Furthermore, when examining the results from the UKF and EKF, around $t = 430$, it is evident that an outlier contributes to the divergence of estimated RUL, leading the filter to provide results with negative amplitude which cannot intersect the EOL threshold. This deviation ultimately impacts the RUL estimation, suggesting an increase towards infinity at that specific time. At the same time, the proposed MCEKF could handle the effect of the outliers and remain in the accuracy bound. It should be noted that for this case, the kernel size of MCEKF was selected to be equal to $\sigma = 0.45$ by trial and error, as in the previous case.

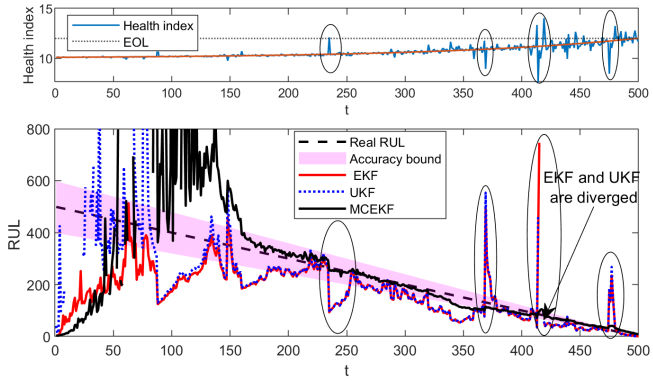


FIGURE 6.7: Final results of predicted RUL in the presence of the non-Gaussian noise, top panel: HI and EOL, bottom panel: the mean of the predicted RUL performed by EKF, UKF, and proposed MCEKF, and the real RUL $\pm 20\%$ accuracy bound.

The PDF portraying the estimated RUL values obtained using the MCEKF technique is shown in Fig. 6.8. As in the previous case, in the initial stages where the dataset contains an insufficient number of data points, the PDF of the estimated RUL does not yield a favorable outcome. However, a discernible improvement becomes apparent as the size of the dataset increases, allowing the PDF to progressively enclose the real RUL. Furthermore, 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 6.9. As can be seen, after $t = 200$, the RUL values are within the scope of the 90% CI. This notable alignment between the real RUL and the CI serves as evidence of the efficacy of proposed model, notably in the context of its robustness against influence of the non-Gaussian noise.

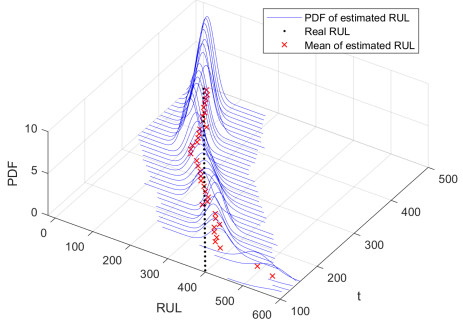


FIGURE 6.8: PDF of the estimated RUL in the presence of the non-Gaussian noise, changing over time.

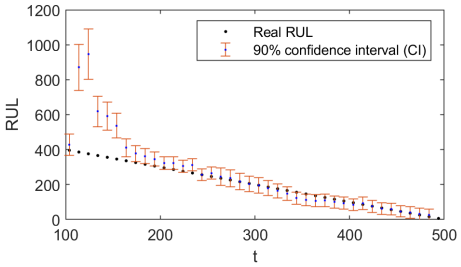


FIGURE 6.9: The 90% CI of the RUL estimated by using MCEKF for simulated HI in the presence of non-Gaussian noise.

To confirm the performance of the proposed method, this procedure was repeated for 100 simulated cases in the presence of Gaussian noise, and the results were presented using box plots (see Fig. 6.10). The MAE was used to compare the performance of the methods; see Eq. (6.8). It should be noted that a lower MAE value indicates better performance. It is defined as follows

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|RUL(t) - \widehat{RUL}(t)\right|, \tag{6.8}$$

where N is the number of all the data points in which the RUL is calculated, $\widehat{RUL}(t)$ is the estimated RUL at the time (data point) t, and $RUL(t) = N - t$ is the real RUL.

As can be seen in Fig. 6.10, the median of the MAE for MCEKF has the lowest value, the same for the first and third quantiles of MCEKF, which means that it worked better to predict RUL in comparison to the other two methods. It should be noted that the kernel size for MCEKF is selected as $\sigma = 0.81$.

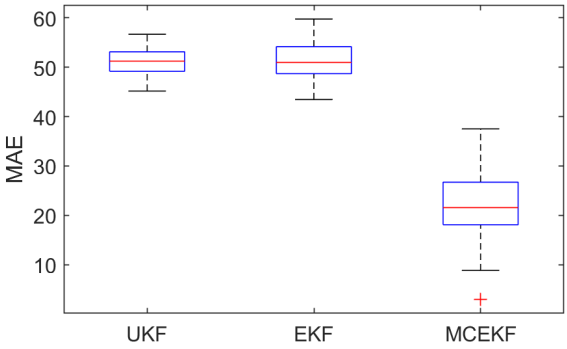


FIGURE 6.10: MAE based on 100 simulated cases in the presence of the Gaussian noise.

To confirm the performance of the proposed method, this procedure was repeated for 100 simulated cases in the presence of non-Gaussian noise (Student's t with $\nu = 2.1$), and the results were presented using boxplots; see Fig. 6.11 as previously. As can be seen in Fig. 6.11, the median of the MAE for MCEKF has the lowest value, the same for the first and third quantiles of MCEKF, which means that this method worked better than the other two to predict RUL. The kernel size for MCEKF is selected as $\sigma = 0.56$.

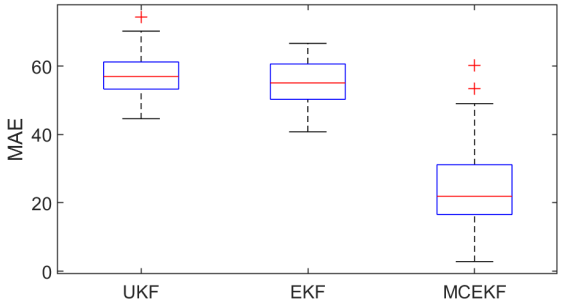


FIGURE 6.11: Boxplots of MAE based on 100 simulated cases in the presence of the non-Gaussian noise (Student's t distributed noise with $\nu = 2.1$).

Furthermore, to clarify the impact of varying levels of non-Gaussianity on the estimated RUL derived from these filters, a simulation involving 100 trajectories for different levels of the degree of freedom ν of Student's t distribution was conducted. Subsequently, these methodologies were used for the estimation of the RUL and the

MAE was calculated by contrasting the actual RUL with the mean of the predicted RUL. The results of these calculated MAE values are illustrated in Fig. 6.12. It is evident from the results that the MCEKF consistently exhibits lower MAE values in different degrees of non-Gaussian noise compared to the outcomes of EKF and UKF. This observation underscores the robustness of MCEKF when faced with non-Gaussian noise scenarios.

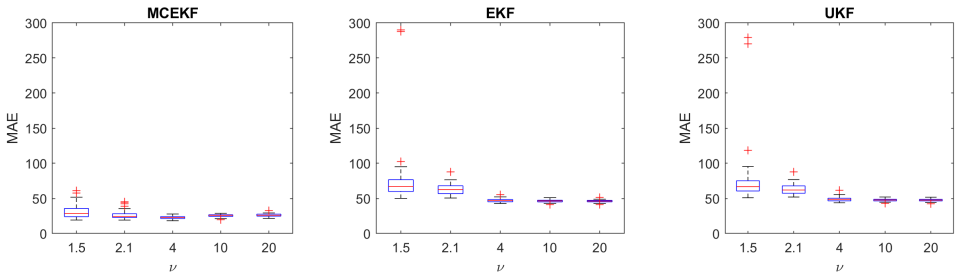


FIGURE 6.12: Boxplots of MAE for 100 simulated trajectories for case of non-Gaussian noise for different levels of non-Gaussianity (Student's t distributed noise with different levels of the ν value). left panel is represented boxplot of MAE for provided result by MCEKF, middle panel is represented boxplot of MAE for provided result by EKF, and right panel is represented boxplot of MAE for provided result by UKF.

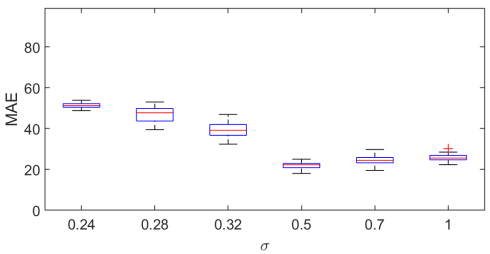


FIGURE 6.13: Boxplot of MAE between real RUL and mean of estimated RUL for 100 simulated trajectories for case of Gaussian noise with different kernel size σ for MCEKF.

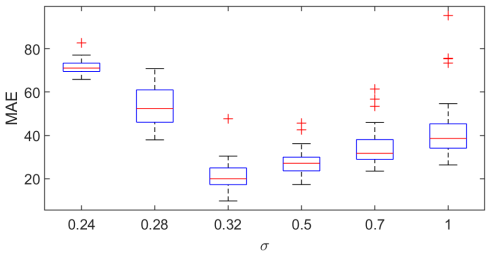


FIGURE 6.14: Boxplot of MAE between real RUL and mean of estimated RUL based for 100 simulated trajectories for case of non-Gaussian noise (Student's t distributed noise with $\nu = 2.1$) with different kernel size σ for MCEKF.

6.3 Result for real dataset

In this section, the results arising from the implementation of the suggested methodology on real-world datasets are explained. Specifically, the efficacy of the approach was investigated in three distinct datasets.

6.3.1 Results for FEMTO dataset

In this subsection, the proposed method is applied to estimate the RUL of the FEMTO dataset from the last regime. The FPT is detected by SMCKF in subsection 5.

Fig. 6.15, shows the results of applying the proposed approach to the FEMTO dataset. Top panel illustrates the degradation curves and the level of EOL. The bottom panel shows the estimated results using EKF, UKF, and the proposed MCEKF. The black dashed line and the purple area show the real RUL value and the $\pm 20\%$ accuracy bound to predict RUL. As could be expected, when the amount of data is lower, the accuracy of the estimated RUL by using all filters is non-desirably low. Based on bottom panel, EKF, UKF and MCEKF first overestimated RUL. After a while, around $t = 30$, it can be seen that all the methods are reaching the accuracy bound and remain approximately inside until the end of their useful life. In addition, the selected parameters for MCEKF are presented in Table 6.1 and the parameters of EKF and UKF are selected as the same values as for MCEKF.

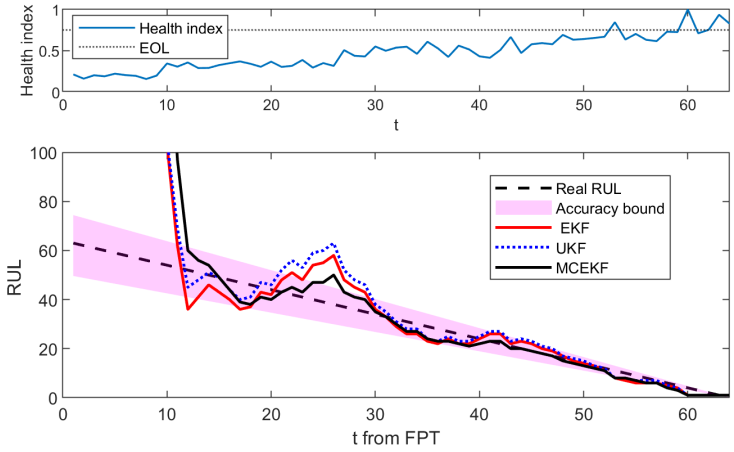


FIGURE 6.15: Final results of estimated RUL for FEMTO dataset, top panel: HI and EOL, bottom panel: predicted RUL results performed by EKF, UKF and proposed MCEKF, and real RUL $\pm 20\%$ accuracy bound.

The PDF showing the estimated RUL values from the MCEKF technique is shown in Fig. 6.16. Evidently, in the initial stages, where the dataset contains an insufficient number of data points, the PDF of the estimated RUL does not yield favorable results. However, a discernible improvement becomes apparent as the dataset size increases, allowing the PDF to progressively enclose the RUL distribution. In particular, beyond the temporal threshold of $t = 30$, the mean of the PDF aligns closely with the actual value of the RUL. Furthermore, 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 6.17. As can be seen; after $t = 30$, the RUL values are within the scope of the 90% CI.

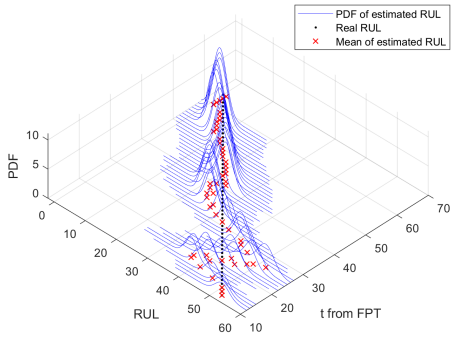


FIGURE 6.16: PDF of estimated RUL, changing over time, obtained by using MCEKF for FEMTO dataset.

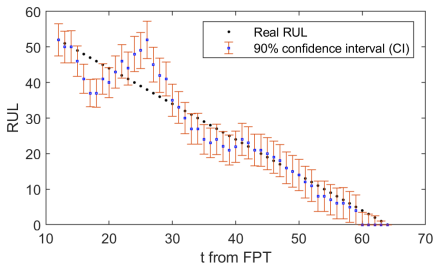


FIGURE 6.17: The 90% CI of estimated RUL by using MCEKF for FEMTO dataset.

To illustrate the sensitivity of the proposed methodology to the hyperparameter σ (kernel bandwidth for the MCEKF), an analysis was conducted employing different ranges of σ. The results related to the estimated RUL are shown in Fig. 6.18. In particular, in this scenario, the selection of a higher value of σ yields more accurate results compared to opting for a lower value of σ. To reinforce this observation, the MAE computed between the real RUL and the estimated RUL using the proposed method is presented in Fig. 6.19, further affirming the findings from the preceding Fig. 6.18.

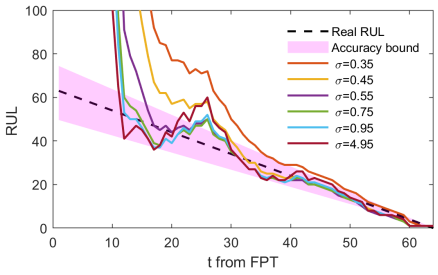


FIGURE 6.18: Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound.

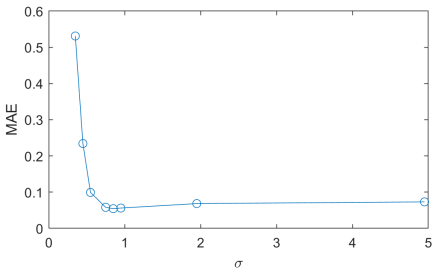


FIGURE 6.19: MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL.

TABLE 6.1: Selected parameters value for MCEKF in case of FEMTO dataset.

MCEKF parameters	Q_t	M_t	σ
FEMTO dataset	diag(0.00001, 0.00001, 0.00001)	0.0038	0.87

6.3.2 Results for IMS dataset

In this part, the proposed method is used to estimate the RUL of IMS dataset from last stage. Panel (a) in Fig. 6.20 shows the complete degradation curves for this case study. Also, the small window shows the area after FPT that is detected by method SMCKF, which is used to estimate RUL.

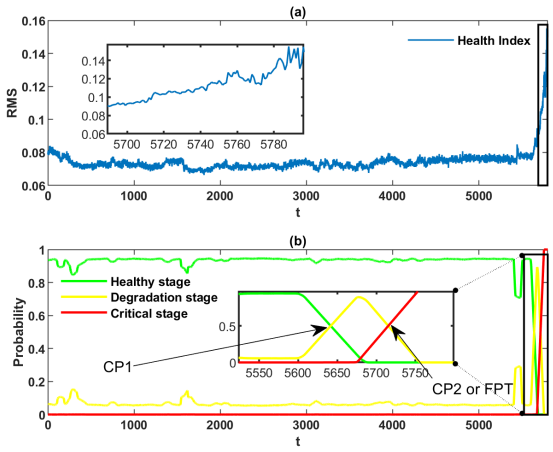


FIGURE 6.20: (a) Health index (RMS) of IMS dataset: subset 3, bearing 3 (b) FPT detected by SMCKF.

The results of applying all of the methods on the IMS dataset are presented in Fig. 6.21. Top panel illustrates the degradation curves and EOL level. Bottom panel shows the estimated results using EKF, UKF, and proposed MCEKF. The black dash line and the purple area show the real RUL value $\pm 20\%$ accuracy bound for predicting RUL. As expected, when the dataset size is small, the accuracy of RUL estimation using all filters is sub-optimal. Based on bottom panel, EKF, UKF, and MCEKF overestimated RUL at first. After a while, around $t = 30$, it can be seen that MCEKF reaches the accuracy bound and remains approximately within it until the end of the life, while UKF and EKF do not provide accurate results. Furthermore, the details of the parameters chosen for the MCEKF are provided in Table 6.2. It is important to note that, for a fair comparative analysis, the parameters of the EKF and the UKF have been set to the same values as those used in the MCEKF.

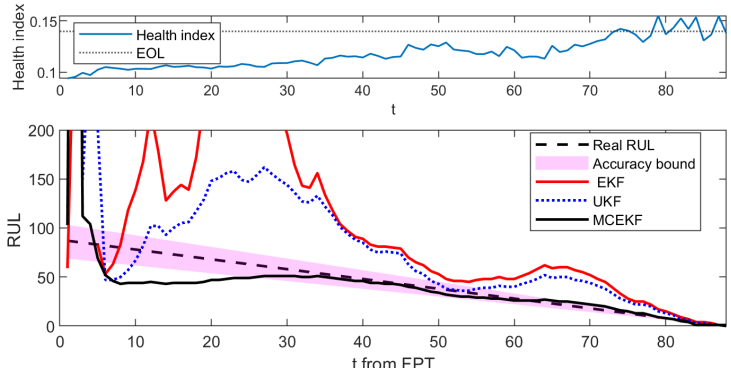


FIGURE 6.21: Final results of estimated RUL for IMS dataset, top panel: HI and EOL, bottom panel: predicted RUL results performed by EKF, UKF and proposed MCEKF, and real RUL $\pm 20\%$ accuracy bound.

The PDF of the estimated RUL values obtained using the MCEKF technique is shown in Fig. 6.22. Evidently, in the initial stages, where the dataset contains an insufficient number of data points, the PDF of the estimated RUL does not yield favorable results. However, a discernible improvement becomes apparent as the dataset size increases, allowing the PDF to progressively enclose the real RUL. In particular, beyond the temporal threshold of t = 30, the mean of the PDF aligns closely with the actual RUL value. Furthermore, 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 6.23. As can be seen; after $t = 30$, the RUL values are within the scope of 90% CI.

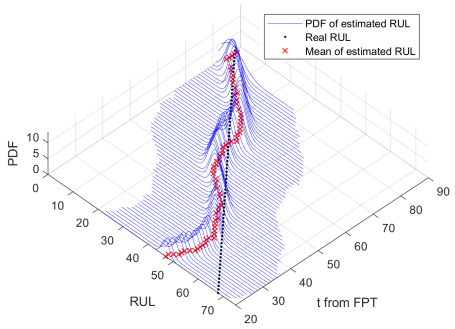


FIGURE 6.22: PDF of estimated RUL, changing over time, obtained by using MCEKF for IMS dataset.

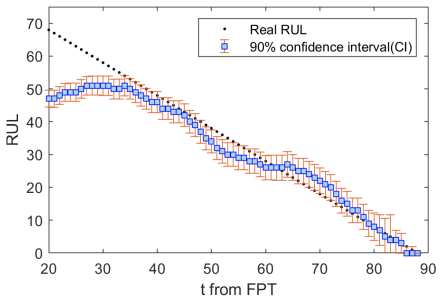


FIGURE 6.23: The 90% CI of estimated RUL by using MCEKF for IMS dataset.

To illustrate the sensitivity of the proposed methodology to the hyperparameter σ (kernel bandwidth for the MCEKF), an analysis was conducted employing different ranges of σ. The results pertaining to the estimated RUL are shown in Fig. 6.24. In particular, in this scenario, the selection of a σ value between 0.25 and 0.4 yields more accurate results compared to opting for a lower or higher value of σ. To reinforce this observation, the MAE calculated between the real RUL and the estimated RUL using the proposed method is presented in Fig. 6.25, further affirming the findings of the preceding Fig. 6.24.

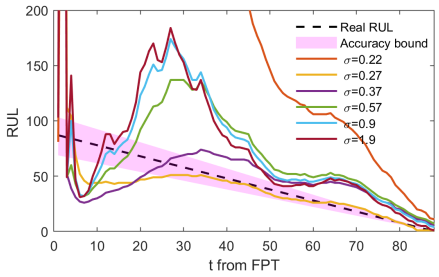


FIGURE 6.24: Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound.

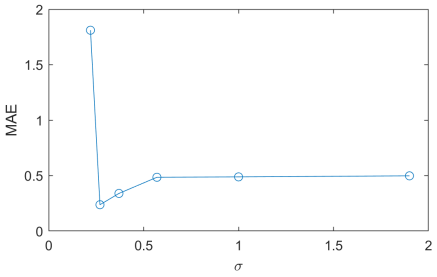


FIGURE 6.25: MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL.

TABLE 6.2: Selected parameters value for MCEKF in case of IMS dataset.

MCEKF parameters	Q_t	M_t	σ
IMS dataset	diag(0.00001, 0.00001, 0.00001)	0.0038	0.27

6.3.3 Results for wind turbine dataset

In this subsection, the proposed method is used to estimate the RUL of the wind turbine dataset. It is important to emphasize that FPT is detected by using SMCKF (see, Subsection 5.2.3). Therefore, the focus here is on the RUL estimation starting from the FPT calculated in previous chapter.

Fig. 6.26, shows the results of applying the proposed method to the wind turbine dataset. Top panel illustrates the degradation curves and the level of EOL. Bottom panel shows the estimated results using EKF, UKF, and the proposed MCEKF. The black dashed line and the purple area show the accuracy bound to the real RUL value $\pm 20\%$ to predict RUL.

As expected, for scenarios where the dataset is limited, the RUL estimates obtained with all filtering methodologies exhibit suboptimal accuracy. The bottom panel analysis shows that the initial results from EKF, UKF, and MCEKF overestimate the RUL. A noticeable convergence towards the accuracy bond can be discerned around $t = 180$. However, it is imperative to acknowledge that due to the non-linear trajectory of the HI, achieving precise results remains a challenge.

Subsequently, the results of all methodologies, obtained iteratively, are shown as time passes. As it can be seen; after $t = 300$, the mean RUL estimated by MCEKF is closely aligned with the accuracy bound, while the other results produced by EKF and UKF exhibit less accurate results. It should be noted that around $t = 310$ and $t = 440$, the RUL estimates derived from EKF and UKF appear to be susceptible to the influence of non-Gaussian noise. In contrast, the MCEKF, as expected, reduced the impact of such non-Gaussian noise. In this study, the parameters of the MCEKF applicable to the case study are carefully chosen by trial and error. These parameter choices are detailed in Table 6.3. Importantly, for a consistent comparison, the measurement noise and process noise parameters have been set to the same values for the EKF, UKF, and MCEKF.

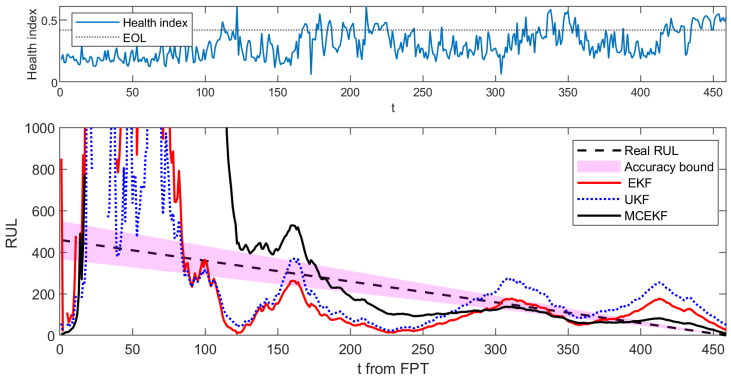


FIGURE 6.26: Final results of estimated RUL for wind turbine dataset, top panel: HI and EOL, bottom panel: predicted RUL results performed by EKF, UKF and proposed MCEKF, and real RUL $\pm 20\%$ accuracy bound.

The PDF of the estimated RUL values from the MCEKF approach is shown in Fig. 6.27. As in previous cases, in the initial stages, where the dataset contains an insufficient number of data points, the PDF of the estimated RUL does not yield favorable outcomes. However, a discernible improvement becomes apparent as the dataset size increases, allowing the PDF to progressively enclose the RUL distribution. In particular, beyond the temporal threshold of $t = 250$, the PDF closely matches the actual RUL value. Moreover, the 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is shown in Fig. 6.28. As can be seen, after $t = 260$, the RUL values are within the scope of the 90% CI. This notable alignment between the real RUL and the 90% CI serves as evidence for the efficacy of the proposed model, particularly in terms of its robustness against non-Gaussian noise influence and the non-linear trend of the HI.

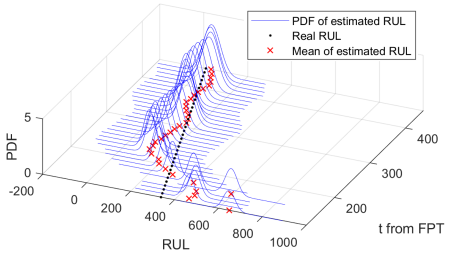


FIGURE 6.27: PDF of estimated RUL, changing over time, obtained by using MCEKF for wind turbine dataset.

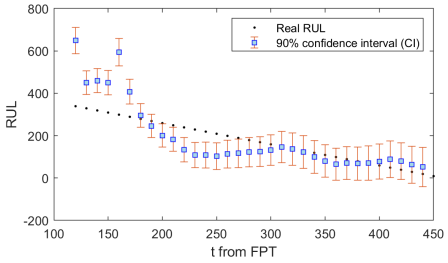


FIGURE 6.28: The 90% CI of estimated RUL by using MCEKF for wind turbine dataset.

To illustrate the sensitivity of the proposed methodology to the hyperparameter σ (kernel bandwidth for the MCEKF), an analysis was conducted using different ranges of σ. The results regarding the estimated RUL are depicted in Fig. 6.29. Specifically, in this scenario, selecting a σ value between 0.25 and 0.4 yields more accurate results compared to choosing either a lower or higher σ value. To underscore this observation, the MAE between the real RUL and the estimated RUL using the proposed method is presented in Fig. 6.30, which further corroborates the findings from Fig. 6.29.

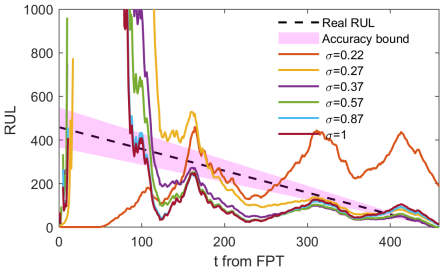


FIGURE 6.29: Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound.

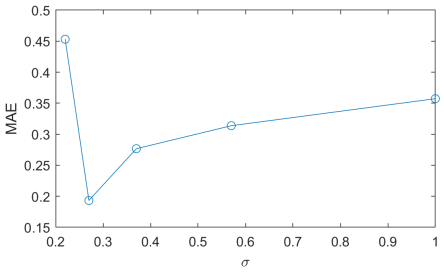


FIGURE 6.30: MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL.

TABLE 6.3: Selected parameters value for MCEKF in case of wind turbine dataset.

MCEKF parameters	Q_t	M_t	σ
wind turbine dataset	diag(0.00001, 0.00001, 0.00001)	0.0265	0.28

6.3.4 Discussion

The effectiveness of the proposed methods has been validated by the results derived from three real-world datasets. The EKF, UKF, and the proposed MCEKF all produce satisfactory and acceptable outcomes for the FEMTO dataset due to its predominantly monotonous nature. Furthermore, earlier research has established that the noise distribution for this particular case study is closely akin to a Gaussian distribution.

The observations from the wind turbine dataset align with expectations and show distinct behavior. The dynamics of this dataset display non-linear characteristics, unlike the monotonic patterns seen in the FEMTO or IMS datasets. This non-monotonic behavior can be attributed to factors such as load fluctuations, self-repair mechanisms, and changes in ambient conditions. The presence of occasional spikes further supports the evidence of a non-Gaussian noise distribution.

However, the proposed method outperforms both the EKF and the UKF for the wind turbine dataset. Despite this, the average predicted RUL produced by all methods fluctuates between overestimation and underestimation as time progresses. It is important to emphasize that the estimation of the RUL is a complex undertaking, encompassing multiple aspects rather than a single prediction point. As demonstrated by the wind turbine dataset, although the mean predicted RUL may differ from the actual RUL, the key point is that the 90% CI effectively encompasses the range of RUL. This validates the strength and effectiveness of the proposed method in dealing with the complex patterns and non-Gaussian noise impact present in the dataset.

To fully understand the performance of the method, the normalized MAE between the actual RUL and the estimated mean RUL in all datasets is provided using Eq. (6.8), as shown in Fig. 6.31. Specifically, it is illustrated in this figure that the proposed MCEKF demonstrates a markedly diminished MAE compared to both the EKF and the UKF.

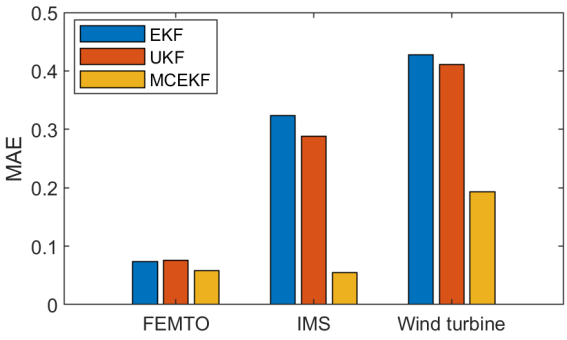


FIGURE 6.31: Calculated MAE for all datasets.

In fact, while the results presented here underscore the substantial potential of the proposed model for the estimation of RUL, the general features of methods from the KF family remain acknowledged. This family of algorithms mainly deals with

the deterministic part of the degradation process, whereas the PDF of the estimated RUL mostly relates to uncertainties in the deterministic part of the degradation process.

In light of this understanding, future endeavors will strategically align with this paradigm. It is proposed to develop a hybrid model that integrates the strengths inherent in the MCEKF with stochastic methodologies designed to specifically encapsulate the stochastic aspects of the degradation process. It is expected that this integration will effectively leverage the precision of MCEKF's deterministic estimation with the power of stochastic methods in modeling the random part, thereby providing a more comprehensive and refined framework for RUL estimation. This approach may have the potential to further enhance the accuracy and robustness of RUL estimation while taking into account both deterministic and stochastic elements.

Chapter 7

Conclusion, contributions and future research

This chapter provides an overall evaluation of the research, presents a concise overview of the contributions made, and proposes potential avenues for further research.

7.1 Conclusion

Nowadays, with the advancement in data acquisition, storage, and transmission technologies, there is a growing emphasis on data-driven approaches for CBM. In industrial settings, vast amounts of data and signals are generated that indicate the condition of machinery. However, these data are often combined with noise and exhibit time-varying behavior, presenting significant challenges in analyzing and utilizing them to predict the future state of machinery for CBM applications. Therefore, in this thesis, the impact of non-Gaussian noise and the non-stationary behavior of HI data are investigated. Solutions are proposed to extract meaningful information from historical HI data, segment the HI, and predict the RUL. The main achievements of this dissertation are summarized below.

1. Development of an artificial HI model for generating HI: Based on observations from real datasets, a model for generating artificial HI data was proposed. This model is based on a three-stage degradation process, which includes healthy, degradation (slow degradation), and critical (fast degradation) states of machinery. Unlike traditional three-stage models that assume constant variance of the random component and Gaussian noise distribution, this thesis extends the model to consider time-varying variance (scale in the case of non-Gaussian noise) and dependencies of the random component, making it more realistic. The proposed model allows for the generation of synthetic data with non-Gaussian noise distributions, such as α-stable and Student's t, providing more flexible tools for simulating realistic HI data.

2. Framework for identifying characteristics of HI: A novel framework for identifying the characteristics of HI based on historical data was developed. This step is crucial before selecting a model for segmentation or prediction, as it involves analyzing and identifying the main characteristics of HI. Proper mathematical tools were proposed to detect the deterministic (global behavior) and random components of HI, which is particularly important when dealing with uncertainty in the presence of non-Gaussian noise. The findings confirmed that non-Gaussian noise significantly affects the proper detection of deterministic and random components, as traditional approaches based on Gaussian assumptions fail under such conditions. The proposed framework, which tolerates the effects of non-Gaussian noise, successfully identifies both deterministic and random components, providing insightful information about the degradation process and forming the basis for developing segmentation and prediction models. The performance of the proposed framework was tested

using simulated signals with Gaussian and different levels of non-Gaussian noise, as well as two real datasets. The first dataset, FEMTO, showed that the identified deterministic part behavior changed from constant to linear and then to exponential or polynomial. The detected variance of the random component exhibited similar behavior. The identified AR coefficients are close to zero, indicating no dependencies in the random part, and fitting distributions on the residuals confirmed that the noise distribution of the dataset was close to Gaussian. In contrast, the wind turbine dataset revealed more complex trends, especially in the random part where the identified scale showed a complex trend at the end of the procedure. Dependencies in the random part were detected, and fitting distributions on the residuals confirmed non-Gaussian noise distribution. The same workflow was repeated using a traditional approach based on Gaussian assumptions. The HI was regenerated based on identified components by both the Gaussian-based framework and the proposed framework. The results confirmed that the Gaussian-based framework could not accurately track the behavior of the original signal in the case of the wind turbine dataset. However, the proposed framework generated HI that closely matched the original wind turbine HI, demonstrating the superiority of the proposed framework over traditional Gaussian-based methods.

3. Development of offline segmentation methods: Based on the findings from the previous step, segmentation methods for historical HI data were developed according to the deterministic part behavior, considering the random noise distribution. Segmentation of historical data is crucial because it allows for localized analysis of HI and enables the development of switching models, where each model could have different behaviors, making complex and precise models particularly for prediction. In this thesis, a segmentation approach based on three polynomial functions (constant, linear, and exponential) was proposed, associated with healthy, degradation, and critical stages, respectively. To demonstrate the effect of non-Gaussian noise, various regression methods, including robust methods and those based on Gaussian assumptions, were employed to fit these three functions to the data and identify the proper change points between stages.

The results for simulated data and three real datasets (FEMTO, IMS, and wind turbine) proved that approaches based on robust methods, such as regression methods using Student's t-loss function or least absolute error loss function, performed better in detecting change points between stages, especially in the presence of non-Gaussian noise. While most methods provided acceptable results in the presence of Gaussian noise, robust methods excelled with non-Gaussian noise. For the FEMTO and IMS datasets, where noise distribution is close to Gaussian, and the wind turbine dataset, where noise distribution is non-Gaussian, the proposed segmentation approach based on dynamic trends (constant, linear, and exponential) was validated. The robust regression methods, particularly those developed using Student's t-loss function, provided the best results for the wind turbine dataset, aligning well with expert analysis. Overall, it was concluded that using robust loss functions is most helpful for segmenting HI into three stages, providing more reliable results.

4. Development of online segmentation methods: Building on the historical data segmentation methods, this thesis developed online segmentation methods to provide probabilistic results about stages in real-time applications. The SMCKF was developed as a robust version of the SKF. This switching approach uses three dynamic models to describe the healthy, degradation, and critical stages, utilizing first, second, and third-order polynomial Kalman filters. Given that the correntropy criterion is more robust than classical mean square error, the proposed SMCKF provided better results compared to the SKF in the presence of non-Gaussian noise.

This robustness was demonstrated by applying the proposed SMCKF to segment simulated HI data in the presence of Gaussian and various levels of non-Gaussian noise. Additionally, the SMCKF and SKF were employed to segment the FEMTO and wind turbine datasets. For the FEMTO dataset, where noise distribution is close to Gaussian, both approaches provided proper results. However, for the wind turbine dataset, where noise distribution is non-Gaussian, the classical SKF was adversely affected by non-Gaussian noise. In contrast, the proposed SMCKF handled the non-Gaussian noise effectively, providing results that better matched the real change points.

Another advantage of this switching model is its Bayesian nature, allowing the integration of historical information extracted from the framework into the model for more accurate results. Additionally, this model provides probabilistic results, valuable for decision-making and health assessment of machinery. The switching model can also perform one-step predictions, useful for machinery health prediction, which will be further investigated in the conclusion.

5. One-step prediction of HI dat for predicting RUL: In the final part of this thesis, a method for one-step prediction of the HI was developed based on an exponential state space degradation model from the first point of the last detected stage to predict RUL. The MCEKF, a robust version of the classical EKF, was applied to the exponential state space degradation model. As previously discussed, the correntropy criterion is more robust than the classical mean square error approach. Thus, the proposed MCEKF approach provided more accurate prediction results compared to the classical EKF.

This was demonstrated by employing the proposed MCEKF approach to predict the next value of the simulated HI in the presence of Gaussian and various levels of non-Gaussian noise. The ultimate validation of the MCEKF method was carried out through its application to three benchmark real-world datasets, including IMS, FEMTO, and wind turbine datasets. This real-world deployment served a dual purpose: to confirm the effectiveness of the proposed approach in accurate HI prediction and to validate its applicability and efficacy within authentic operational environments.

The introduced methods demonstrated superiority through comparative simulations, coupled with its successful deployment on real-world datasets, confirms its potential as an invaluable tool in the realm of mechanical systems prognostics and health management. Additionally, based on the Bayesian nature of MCEKF, it is possible to utilize historical data gained from the framework to provide more accurate results. The probabilistic results are also useful for dealing with uncertainties. Furthermore, by providing a threshold for the end of life, it is possible to RUL.

In conclusion, this research provides a robust framework for handling long-term degradation data amidst non-Gaussian noise, with significant implications for machine learning, condition monitoring, and predictive maintenance. The models and methods developed in this thesis offer a comprehensive solution to the challenges in HI data analysis and prognostics, improving the accuracy and reliability of degradation assessments and predictions.

7.2 Practical perspectives on the proposed approaches

In industries, like mining industry, where data acquisition is plagued by heavy-tailed noise, practical application of the theories and methodologies discussed in this

thesis could significantly enhance CBM strategies. Mining operations typically generate vast amounts of data from machinery and sensors that are intrinsically noisy and non-Gaussian in nature. The development of an artificial HI model that accounts for non-Gaussian noise distributions, such as α-stable and Student's t, provides a more realistic simulation of actual HI data, reflecting true operational environments. Furthermore, the novel framework for identifying characteristics of HI can be particularly beneficial in recognizing patterns, anomalies of machines. Segmentation of HI to different degradation stages despite the noise, which is crucial for timely maintenance decisions. Employing segmentation methods that consider non-Gaussian noise, and the subsequent robust probablstic techniques like SMCKF, can significantly improve the accuracy in determining the critical change points between different machinery states. These methodologies enable mining companies to better predict machinery failures and optimize maintenance schedules, thus reducing downtime and enhancing operational efficiency. This practical approach underscores the importance of adapting advanced data analysis techniques to the specific challenges posed by industries with heavy-tailed noise in their data streams.

7.3 Contributions

This PhD thesis makes several significant contributions to the field of long-term HI degradation data analysis, particularly in the presence of non-Gaussian noise:

- A novel three-stage model, incorporating time-varying characteristics and non-Gaussian noise in the random component, has been developed for generating long-term HI degradation data.
- A new framework for robustly identifying the characteristics of long-term HI data, considering non-Gaussian noise, has been developed. This framework can be applied to real HI datasets to precisely analyze them.
- A novel switching method based on the SMCKF based on the three-stage model has been derived to identify each stage of the degradation process while accounting for non-Gaussian noise. This model can be used for probabilistic and online machinery health assessments.
- A new method based the MCEKF, based on a state-space exponential degradation model, has been developed to provide probabilistic results for estimating RUL while considering non-Gaussian noise.

7.4 Future research

This thesis opens several avenues for future research in the field of long-term HI degradation data analysis, particularly in the presence of non-Gaussian noise. Some promising topics for further investigation include:

- Developing advanced machine learning algorithms to handle non-Gaussian noise in HI data more effectively.
- Exploring the use of deep learning techniques for prognostics and health assessments in the presence of non-Gaussian noise.
- Studying the development of optimization algorithms for tuning model parameters in the presence of non-Gaussian noise to improve the accuracy of health assessments.

- Exploring the integration of domain knowledge and expert systems to enhance the performance of health assessment models in scenarios with non-Gaussian noise.

Appendix A

Considered distributions

- The Gaussian distribution $\mathcal{N}(\mu, \sigma)$ is defined through the PDF [154, 126]

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}, \tag{A.1}$$

  where $\mu \in \mathbb{R}$ is the mean and $\sigma > 0$ is the standard deviation.

- The α-stable distribution $\mathcal{S}(\alpha, \beta, \gamma, \delta)$ is defined through the characteristic function [155, 156]

$$\Phi(t) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left[1 + i\beta \operatorname{sign}(t) \tan\frac{\pi\alpha}{2}\left((\gamma|t|)^{1-\alpha} - 1\right)\right] + i\delta t\right) & \text{for } \alpha \neq 1, \\ \exp\left(-\gamma|t| \left[1 + i\beta \operatorname{sign}(t) \frac{2}{\pi} \ln(\gamma|t|)\right] + i\delta t\right) & \text{for } \alpha = 1, \end{cases} \tag{A.2}$$

  where $0 < \alpha \leq 2$ - stability parameter, $-1 \leq \beta \leq 1$ - skewness parameter, $\gamma > 0$ - scale parameter and $\delta \in \mathbb{R}$ - location parameter. Notice that Gaussian distribution is a special case of the α-stable distribution. The α-stable distribution with $\alpha = 2$ corresponds to the normal distribution. It is the only case when the variance is finite.

- The Student's t distribution is a fundamental statistical tool that incorporates an essential parameter: degrees of freedom ν. This distribution provides a robust framework for modeling data, accommodating deviations from normality. Widely utilized in statistical inference, hypothesis testing, and modeling scenarios involving small sample sizes or unknown population standard deviation, the Student's t distribution enhances the accuracy and reliability of statistical conclusions.

  The probability density function of the Student's t distribution is expressed as follows:

$$f(x; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{A.3}$$

  where Γ is Gama function.The mean of the Student's t distribution exists if $\nu > 1$. For $\nu \leq 1$, the mean is undefined. The variance exists if $\nu > 2$. For $1 < \nu \leq 2$, the variance is infinite, and for $\nu \leq 1$, the variance is undefined.

- The t location-scale distribution $\mathbf{t}(\mu, \sigma, \nu)$ is defined through the PDF [157, 158]

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left[\frac{\nu + \left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{\left(\frac{\nu+1}{2}\right)}, \tag{A.4}$$

where , $\mu \in \mathbb{R}$ - location parameter, $\sigma > 0$ - scale parameter and $\nu > 0$ - shape parameter. The variance is only defined for values of $\nu > 2$. If a given random variable $X \sim \mathbf{t}(\mu, \sigma, \nu)$, then $(X - \mu)/\sigma$ has a Student's t distribution with ν degrees of freedom. When ν tends to infinity, then Student's t distribution tends to a Gaussian distribution.

Bibliography

[1] EN Cen. BSI Standard Publication: BS EN 13306:2010 Maintenance", Maintenance terminology. 2010.

[2] Kevin Swearingen, Wayne Majkowski, Brian Bruggeman, Dan Gilbertson, Jon Dunsdon, and Ben Sykes. "An open system architecture for condition based maintenance overview". In: 2007 IEEE Aerospace Conference. IEEE. 2007, pp. 1–8.

[3] J Gordon Leishman. Introduction to Aerospace Flight Vehicles. Embry-Riddle Aeronautical University, 2023.

[4] Alberto Diez-Olivan, Javier Del Ser, Diego Galar, and Basilio Sierra. "Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0". In: Information Fusion 50 (2019), pp. 92–111.

[5] Forough Moosavi, Hamid Shiri, Jacek Wodecki, Agnieszka Wyłomańska, and Radoslaw Zimroz. "Application of Machine Learning Tools for Long-Term Diagnostic Feature Data Segmentation". In: Applied Sciences 12.13 (2022), p. 6766.

[6] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou. "Remaining useful life estimation–a review on the statistical data driven approaches". In: European journal of operational research 213.1 (2011), pp. 1–14.

[7] Zhi-Sheng Ye and Min Xie. "Stochastic modelling and analysis of degradation for highly reliable products". In: Applied Stochastic Models in Business and Industry 31.1 (2015), pp. 16–32.

[8] Aiwina Heng, Sheng Zhang, Andy CC Tan, and Joseph Mathew. "Rotating machinery prognostics: State of the art, challenges and opportunities". In: Mechanical Systems and Signal Processing 23.3 (2009), pp. 724–739.

[9] Linxia Liao and Felix Köttig. "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction". In: IEEE Transactions on Reliability 63.1 (2014), pp. 191–207.

[10] Zhibin Zhao, Jingyao Wu, Tianfu Li, Chuang Sun, Ruqiang Yan, and Xuefeng Chen. "Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review". In: Chinese Journal of Mechanical Engineering 34.1 (2021), pp. 1–29.

[11] Paul Paris and Fazil Erdogan. "A critical analysis of crack propagation laws". In: (1963).

[12] Yuning Qian, Ruqiang Yan, and Robert X Gao. "A multi-time scale approach to remaining useful life prediction in rolling bearing". In: Mechanical Systems and Signal Processing 83 (2017), pp. 549–567.

[13] Dong Xu, Jin'e Huang, Qin Zhu, Xun Chen, Yongcheng Xu, and Shuang Wang. "Residual fatigue life prediction of ball bearings based on Paris law and RMS". In: Chinese Journal of Mechanical Engineering 25.2 (2012), pp. 320–327.

[14] Jinjiang Wang, Robert X Gao, Zhuang Yuan, Zhaoyan Fan, and Laibin Zhang. "A joint particle filter and expectation maximization approach to machine condition prognosis". In: Journal of Intelligent Manufacturing 30 (2019), pp. 605–621.

[15] Yaguo Lei, Naipeng Li, Szymon Gontarz, Jing Lin, Stanislaw Radkowski, and Jacek Dybala. "A model-based method for remaining useful life prediction of machinery". In: IEEE Transactions on reliability 65.3 (2016), pp. 1314–1326.

[16] Linxia Liao. "Discovering prognostic features using genetic programming in remaining useful life prediction". In: IEEE Transactions on Industrial Electronics 61.5 (2013), pp. 2464–2472.

[17] Jianzhong Sun, Hongfu Zuo, Wenbin Wang, and Michael G Pecht. "Prognostics uncertainty reduction by fusing on-line monitoring data based on a state-space-based degradation model". In: Mechanical Systems and Signal Processing 45.2 (2014), pp. 396–407.

[18] Charles H Oppenheimer and Kenneth A Loparo. "Physically based diagnosis and prognosis of cracked rotor shafts". In: Component and systems diagnostics, prognostics, and health management II. Vol. 4733. SPIE. 2002, pp. 122–132.

[19] WM Ostachowicz and M Krawczuk. "Coupled torsional and bending vibrations of a rotor with an open crack". In: Archive of applied mechanics 62.3 (1992), pp. 191–201.

[20] Piero Baraldi, Francesca Mangili, and Enrico Zio. "Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data". In: Reliability Engineering & System Safety 112 (2013), pp. 94–108.

[21] Yang Hu, Piero Baraldi, Francesco Di Maio, and Enrico Zio. "Online performance assessment method for a model-based prognostic approach". In: IEEE Transactions on reliability 65.2 (2015), pp. 718–735.

[22] Kwai S Chan, Michael P Enright, Jonathan P Moody, Benjamin Hocking, and Simeon HK Fitch. "Life prediction for turbopropulsion systems under dwell fatigue conditions". In: Journal of engineering for gas turbines and power 134.12 (2012), p. 122501.

[23] Khaled El-Tawil and Abdo Abou Jaoude. "Stochastic and nonlinear-based prognostic model". In: Systems Science & Control Engineering: An Open Access Journal 1.1 (2013), pp. 66–81.

[24] Yudong Cao, Yifei Ding, Minping Jia, and Rushuai Tian. "A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings". In: Reliability Engineering & System Safety 215 (2021).

[25] Zunya Shi and Abdallah Chehade. "A dual-LSTM framework combining change point detection and remaining useful life prediction". In: Reliability Engineering & System Safety 205 (2021).

[26] Bingxin Yan, Xiaobing Ma, Li Yang, Han Wang, and Tianyi Wu. "A novel degradation-rate-volatility related effect Wiener process model with its extension to accelerated ageing data analysis". In: Reliability Engineering & System Safety 204 (2020).

[27] Khanh Le Son, Mitra Fouladirad, and Anne Barros. "Remaining useful lifetime estimation and noisy gamma deterioration process". In: Reliability Engineering & System Safety 149 (2016), 76 – 87.

[28] Khanh T.P. Nguyen, Mitra Fouladirad, and Antoine Grall. "Model selection for degradation modeling and prognosis with health monitoring data". In: Reliability Engineering & System Safety 169 (2018), 105 – 116.

[29] Yaguo Lei, Naipeng Li, Liang Guo, Ningbo Li, Tao Yan, and Jing Lin. "Machinery health prognostics: A systematic review from data acquisition to RUL prediction". In: Mechanical systems and signal processing 104 (2018), pp. 799–834.

[30] Claudio Sbarufatti, Matteo Corbetta, Andrea Manes, and Marco Giglio. "Sequential Monte-Carlo sampling based on a committee of artificial neural networks for posterior state estimation and residual lifetime prediction". In: International Journal of Fatigue 83 (2016), pp. 10–23.

[31] Yongping Pan, Meng Joo Er, Xiang Li, Haoyong Yu, and Rafael Gouriveau. "Machine health condition prediction via online dynamic fuzzy neural networks". In: Engineering Applications of Artificial Intelligence 35 (2014), pp. 105–113.

[32] Lei Xiao, Xiaohui Chen, Xinghui Zhang, and Min Liu. "A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition". In: Journal of Intelligent Manufacturing 28.8 (2017), pp. 1893–1914.

[33] Ryad Zemouri, Daniel Racoceanu, and Noureddine Zerhouni. "Recurrent radial basis function network for time-series prediction". In: Engineering Applications of Artificial Intelligence 16.5-6 (2003), pp. 453–463.

[34] Arnaz Malhi, Ruqiang Yan, and Robert X Gao. "Prognosis of defect propagation based on recurrent neural networks". In: IEEE Transactions on Instrumentation and Measurement 60.3 (2011), pp. 703–711.

[35] Mohamed El-Koujok, Rafael Gouriveau, and Noureddine Zerhouni. "Reducing arbitrary choices in model building for prognostics: An approach by applying parsimony principle on an evolving neuro-fuzzy system". In: Microelectronics reliability 51.2 (2011), pp. 310–320.

[36] Rogério Ishibashi and Cairo Lúcio Nascimento Júnior. "GFRBS-PHM: A genetic fuzzy rule-based system for phm with improved interpretability". In: 2013 IEEE conference on prognostics and health management (PHM). IEEE. 2013, pp. 1–7.

[37] Datong Liu, Jingyue Pang, Jianbao Zhou, Yu Peng, and Michael Pecht. "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression". In: Microelectronics Reliability 53.6 (2013), pp. 832–839.

[38] Sylvester A Aye and PS Heyns. "An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission". In: Mechanical Systems and Signal Processing 84 (2017), pp. 485–498.

[39] Yaguo Lei, Naipeng Li, and Jing Lin. "A new method based on stochastic process models for machine remaining useful life prediction". In: IEEE Transactions on Instrumentation and Measurement 65.12 (2016), pp. 2671–2684.

[40] Joanna Z Sikorska, Melinda Hodkiewicz, and Lin Ma. "Prognostic modelling options for remaining useful life estimation by industry". In: Mechanical systems and signal processing 25.5 (2011), pp. 1803–1836.

[41] Yuning Qian, Ruqiang Yan, and Shijie Hu. "Bearing degradation evaluation using recurrence quantification analysis and Kalman filter". In: IEEE Transactions on Instrumentation and Measurement 63.11 (2014), pp. 2599–2610.

[42] Diana Barraza-Barraza, Víctor G Tercero-Gómez, Mario G Beruvides, and Jorge Limón-Robles. "An adaptive ARX model to estimate the RUL of aluminum plates based on its crack growth". In: Mechanical Systems and Signal Processing 82 (2017), pp. 519–536.

[43] Teresa Escobet, Joseba Quevedo, and Vicenç Puig. "A fault/anomaly system prognosis using a data-driven approach considering uncertainty". In: The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE. 2012, pp. 1–7.

[44] Linkan Bian and Nagi Gebraeel. "Stochastic methodology for prognostics under continuously varying environmental profiles". In: Statistical Analysis and Data Mining: The ASA Data Science Journal 6.3 (2013), pp. 260–270.

[45] Kjell A Doksum and Arnljot Hbyland. "Models for variable-stress accelerated life testing experiments based on wener processes and the inverse gaussian distribution". In: Technometrics 34.1 (1992), pp. 74–82.

[46] George Alex Whitmore and Fred Schenkelberg. "Modelling accelerated degradation data using Wiener diffusion with a time scale transformation". In: Lifetime data analysis 3 (1997), pp. 27–45.

[47] Sheng-Tsaing Tseng and Chien-Yu Peng. "Optimal burn-in policy by using an integrated Wiener process". In: Iie Transactions 36.12 (2004), pp. 1161–1170.

[48] Chanseok Park and William J Padgett. "Stochastic degradation models with several accelerating variables". In: IEEE Transactions on Reliability 55.2 (2006), pp. 379–390.

[49] Nagi Z Gebraeel, Mark A Lawley, Rong Li, and Jennifer K Ryan. "Residual-life distributions from component degradation signals: A Bayesian approach". In: IiE Transactions 37.6 (2005), pp. 543–557.

[50] Xiaopeng Xi, Maoyin Chen, and Donghua Zhou. "Remaining useful life prediction for degradation processes with memory effects". In: IEEE Transactions on Reliability 66.3 (2017), pp. 751–760.

[51] Hanwen Zhang, Chao Jia, and Maoyin Chen. "Remaining useful life prediction for degradation processes with dependent and nonstationary increments". In: IEEE Transactions on Instrumentation and Measurement 70 (2021), pp. 1–12.

[52] C Joseph Lu and William O Meeker. "Using degradation measures to estimate a time-to-failure distribution". In: Technometrics 35.2 (1993), pp. 161–174.

[53] William Q Meeker, Luis A Escobar, and C Joseph Lu. "Accelerated degradation tests: modeling and analysis". In: Technometrics 40.2 (1998), pp. 89–99.

[54] Jong In Park and Suk Joo Bae. "Direct prediction methods on lifetime distribution of organic light-emitting diodes from accelerated degradation tests". In: IEEE Transactions on Reliability 59.1 (2010), pp. 74–90.

[55] Jamie Coble and J WESLEY HINES. "Incorporating prior belief in the general path model: A comparison of information sources". In: Nuclear Engineering and Technology 46.6 (2014), pp. 773–782.

[56] Qiang Li, Zhenhui Ma, Hongkun Li, Xuejun Liu, Xichun Guan, and Peihua Tian. "Remaining useful life prediction of mechanical system based on performance evaluation and geometric fractional Lévy stable motion with adaptive nonlinear drift". In: Mechanical Systems and Signal Processing 184 (2023), p. 109679.

[57] Jan M Van Noortwijk. "A survey of the application of gamma processes in maintenance". In: Reliability Engineering & System Safety 94.1 (2009), pp. 2–21.

[58] Sebastian P Kuniewski, Johannes AM van der Weide, and Jan M van Noortwijk. "Sampling inspection for the evaluation of time-dependent reliability of deteriorating systems under imperfect defect detection". In: Reliability Engineering & System Safety 94.9 (2009), pp. 1480–1490.

[59] Vilijandas Bagdonavicius and Mikhail S Nikulin. "Estimation in degradation models with explanatory variables". In: Lifetime Data Analysis 7 (2001), pp. 85–103.

[60] Jerry Lawless and Martin Crowder. "Covariates and random effects in a gamma process model with application to degradation and failure". In: Lifetime data analysis 10 (2004), pp. 213–227.

[61] He Liu, Wanqing Song, Yuhui Niu, and Enrico Zio. "A generalized cauchy method for remaining useful life prediction of wind turbine gearboxes". In: Mechanical Systems and Signal Processing 153 (2021), p. 107471.

[62] Guangxu Hong, Wanqing Song, Yan Gao, Enrico Zio, and Aleksey Kudreyko. "An iterative model of the generalized Cauchy process for predicting the remaining useful life of lithium-ion batteries". In: Measurement 187 (2022), p. 110269.

[63] Shouwu Duan, Wanqing Song, Enrico Zio, Carlo Cattani, and Ming Li. "Product technical life prediction based on multi-modes and fractional Lévy stable motion". In: Mechanical Systems and Signal Processing 161 (2021), p. 107974.

[64] David E Acuña and Marcos E Orchard. "Particle-filtering-based failure prognosis via sigma-points: Application to lithium-ion battery state-of-charge monitoring". In: Mechanical Systems and Signal Processing 85 (2017), pp. 827–848.

[65] Jiuping Xu, Yusheng Wang, and Lei Xu. "PHM-oriented integrated fusion prognostics for aircraft engines based on sensor data". In: IEEE Sensors Journal 14.4 (2013), pp. 1124–1132.

[66] Francesco Di Maio, Kwok Leung Tsui, and Enrico Zio. "Combining relevance vector machines and exponential regression for bearing residual life estimation". In: Mechanical systems and signal processing 31 (2012), pp. 405–427.

[67] Chaitanya Sankavaram, Bharath Pattipati, Anuradha Kodali, Krishna Pattipati, Mohammad Azam, Sachin Kumar, and Michael Pecht. "Model-based and data-driven prognosis of automotive and electronic systems". In: 2009 IEEE international conference on automation science and engineering. IEEE. 2009, pp. 96–101.

[68] Yumei Ye, Qiang Yang, Jingang Zhang, Songhe Meng, and Jun Wang. "A dynamic data driven reliability prognosis method for structural digital twin and experimental validation". In: Reliability Engineering & System Safety 240 (2023).

[69] Samaneh Aminikhanghahi and Diane J Cook. "A survey of methods for time series change point detection". In: Knowledge and information systems 51.2 (2017), pp. 339–367.

[70] Manfred Mudelsee. "Trend analysis of climate time series: A review of methods". In: Earth-science reviews 190 (2019), pp. 310–322.

[71] Bo Zhou and Hao Ye. "A study of polynomial fit-based methods for qualitative trend analysis". In: Journal of Process Control 37 (2016), pp. 21–33.

[72] Allou Samé, Faicel Chamroukhi, Gérard Govaert, and Patrice Aknin. "Model-based clustering and segmentation of time series with changes in regime". In: Advances in Data Analysis and Classification 5.4 (2011), pp. 301–321.

[73] Seungmin Lee, Yoosoo Jeong, Daejin Park, Byoung-Ju Yun, and Kil Houm Park. "Efficient fiducial point detection of ECG QRS complex based on polygonal approximation". In: Sensors 18.12 (2018), p. 4502.

[74] Amin Shirvani, Klaus Arpe, and Marzieh Jahandideh. "Analysis of trends and change points in meteorological variables over the south of the Caspian Sea". In: Theoretical and Applied Climatology 141.3 (2020), pp. 959–966.

[75] Karolina Gąsior, Hanna Urbańska, Aleksandra Grzesiek, Radoslaw Zimroz, and Agnieszka Wyłomańska. "Identification, decomposition and segmentation of impulsive vibration signals with deterministic components—a sieving screen case study". In: Sensors (Switzerland) 20.19 (2020), 1 – 20.

[76] Daniel Kucharczyk, Agnieszka Wyłomańska, Jakub Obuchowski, Radoslaw Zimroz, and Maciej Madziarz. "Stochastic Modelling as a Tool for Seismic Signals Segmentation". In: Shock and Vibration 2016 (2016).

[77] Aleksandra Grzesiek, Karolina Gasior, Agnieszka Wyłomańska, and Radoslaw Zimroz. "Divergence-based segmentation algorithm for heavy-tailed acoustic signals with time-varying characteristics". In: Sensors 21.24 (2021).

[78] Alkan Alkaya and İlyas Eker. "Variance sensitive adaptive threshold-based PCA method for fault detection with experimental application". In: ISA transactions 50.2 (2011), pp. 287–302.

[79] Olga Fink, Enrico Zio, and Ulrich Weidmann. "A classification framework for predicting components' remaining useful life based on discrete-event diagnostic data". In: IEEE Transactions on Reliability 64.3 (2015), pp. 1049–1056.

[80] Lei Hu, Niaoqing Hu, Xinpeng Zhang, Fengshou Gu, and Ming Gao. "Novelty detection methods for online health monitoring and post data analysis of turbopumps". In: Journal of Mechanical Science and Technology 27.7 (2013), pp. 1933–1942.

[81] James Kuria Kimotho, Christopher Sondermann-Wölke, Tobias Meyer, and Walter Sextro. "Machinery Prognostic Method Based on Multi-Class Support Vector Machines and Hybrid Differential Evolution–Particle Swarm Optimization". In: Chemical Engineering Transactions 33 (2013).

[82] Edwin Sutrisno, Hyunseok Oh, Arvind Sai Sarathi Vasan, and Michael Pecht. "Estimation of remaining useful life of ball bearings using data driven methodologies". In: 2012 ieee conference on prognostics and health management. IEEE. 2012, pp. 1–7.

[83] Yaogang Hu, Hui Li, Xinglin Liao, Erbing Song, Haitao Liu, and Zhe Chen. "A probability evaluation method of early deterioration condition for the critical components of wind turbine generator systems". In: Mechanical Systems and Signal Processing 76 (2016), pp. 729–741.

[84] Prasanna Tamilselvan, Yibin Wang, and Pingfeng Wang. "Deep belief network based state classification for structural health diagnosis". In: 2012 IEEE Aerospace Conference. IEEE. 2012, pp. 1–11.

[85] Junqiang Liu, Fan Lei, Chunlu Pan, Dongbin Hu, and Hongfu Zuo. "Prediction of remaining useful life of multi-stage aero-engine based on clustering and LSTM fusion". In: Reliability Engineering & System Safety 214 (2021), p. 107807.

[86] Jaskaran Singh, AK Darpe, and Satinder P Singh. "Bearing remaining useful life estimation using an adaptive data-driven model based on health state change point identification and K-means clustering". In: Measurement Science and Technology 31.8 (2020), p. 085601.

[87] Wentao Mao, Jianliang He, Bin Sun, and Liyun Wang. "Prediction of Bearings Remaining Useful Life Across Working Conditions Based on Transfer Learning and Time Series Clustering". In: IEEE Access 9 (2021), pp. 135285–135303.

[88] Andrea Giantomassi, Francesco Ferracuti, Alessandro Benini, Gianluca Ippoliti, Sauro Longhi, and Antonio Petrucci. "Hidden Markov model for health estimation and prognosis of turbofan engines". In: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Vol. 54808. 2011, pp. 681–689.

[89] F Sloukia, Mohamed El Aroussi, Hicham Medromi, and Mohamed Wahbi. "Bearings prognostic using mixture of Gaussians hidden markov model and support vector machine". In: 2013 ACS International Conference on Computer Systems and Applications (AICCSA). IEEE. 2013, pp. 1–4.

[90] Joanna Janczura, Tomasz Barszcz, Radosław Zimroz, and Agnieszka Wyłomańska. "Machine condition change detection based on data segmentation using a three-regime, α-stable Hidden Markov Model". In: Measurement 220 (2023), p. 113399.

[91] Chi Keong Reuben Lim and David Mba. "Switching Kalman filter for failure prognostic". In: Mechanical Systems and Signal Processing 52 (2015), pp. 426–435.

[92] Hamid Shiri, Pawel Zimroz, Jacek Wodecki, Agnieszka Wyłomańska, Radosław Zimroz, and Krzysztof Szabat. "Using long-term condition monitoring data with non-Gaussian noise for online diagnostics". In: Mechanical Systems and Signal Processing 200 (2023), p. 110472.

[93] Lingli Cui, Xin Wang, Yonggang Xu, Hong Jiang, and Jianping Zhou. "A novel switching unscented Kalman filter method for remaining useful life prediction of rolling bearing". In: Measurement 135 (2019), pp. 678–684.

[94] Jacek Wodecki, Anna Michalak, and Radosław Zimroz. "Local damage detection based on vibration data analysis in the presence of Gaussian and heavy-tailed impulsive noise". In: Measurement 169 (2021), p. 108400.

[95] Justyna Hebda-Sobkowicz, Jakub Nowicki, Radosław Zimroz, and Agnieszka Wyłomańska. "Alternative Measures of Dependence for Cyclic Behaviour Identification in the Signal with Impulsive Noise—Application to the Local Damage Detection". In: Electronics 10.15 (2021), p. 1863.

[96] Piotr Kruczek, Radosław Zimroz, and Agnieszka Wyłomańska. "How to detect the cyclostationarity in heavy-tailed distributed signals". In: Signal Processing 172 (2020), p. 107514.

[97] Patrick Nectoux, Rafael Gouriveau, Emmanuel Medjaher, Brigitte Chebel-Morello, Noureddine Zerhouni, and Christophe Varnier. "PRONOSTIA: An experimental platform for bearings accelerated degradation tests." In: IEEE International Conference on Prognostics and Health Management, PHM'12. IEEE Catalog Number: CPF12PHM-CDR. 2012, pp. 1–8.

[98] Ahmed Mosallam, Kamal Medjaher, and Noureddine Zerhouni. "Time series trending for condition assessment and prognostics". In: Journal of manufacturing technology management (2014).

[99] Theodoros H Loutas, Dimitrios Roulias, and George Georgoulas. "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression". In: IEEE Transactions on Reliability 62.4 (2013), pp. 821–832.

[100] Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni, and Patrick Nectoux. "Enabling health monitoring approach based on vibration data for accurate prognostics". In: IEEE Transactions on industrial electronics 62.1 (2014), pp. 647–656.

[101] Rodney K Singleton, Elias G Strangas, and Selin Aviyente. "Extended Kalman filtering for remaining-useful-life estimation of bearings". In: IEEE Transactions on Industrial Electronics 62.3 (2014), pp. 1781–1790.

[102] Bin Zhang, Lijun Zhang, and Jinwu Xu. "Degradation feature selection for remaining useful life prediction of rolling element bearings". In: Quality and Reliability Engineering International 32.2 (2016), pp. 547–554.

[103] Sheng Hong, Zheng Zhou, Enrico Zio, and Wenbin Wang. "An adaptive method for health trend prediction of rotating bearings". In: Digital Signal Processing 35 (2014), pp. 117–123.

[104] Yuting Nie and Jiuqing Wan. "Estimation of remaining useful life of bearings using sparse representation method". In: 2015 Prognostics and System Health Management Conference (PHM). IEEE. 2015, pp. 1–6.

[105] Zhiliang Liu, Ming J Zuo, and Yong Qin. "Remaining useful life prediction of rolling element bearings based on health state assessment". In: Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 230.2 (2016), pp. 314–330.

[106] Daniel Zurita, Jesús A Carino, Miguel Delgado, and Juan A Ortega. "Distributed neuro-fuzzy feature forecasting approach for condition monitoring". In: Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA). IEEE. 2014, pp. 1–8.

[107] Liang Guo, Hongli Gao, Haifeng Huang, Xiang He, and ShiChao Li. "Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring". In: Shock and Vibration 2016 (2016).

[108] Xiaohang Jin, Yi Sun, Zijun Que, Yu Wang, and Tommy WS Chow. "Anomaly detection and fault prognosis for bearings". In: IEEE Transactions on Instrumentation and Measurement 65.9 (2016), pp. 2046–2054.

[109] Hongkun Li and Yinhu Wang. "Rolling bearing reliability estimation based on logistic regression model". In: 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE). IEEE. 2013, pp. 1730–1733.

[110] Zeyi Huang, Zhengguo Xu, Xiaojie Ke, Wenhai Wang, and Youxian Sun. "Remaining useful life prediction for an adaptive skew-Wiener process model". In: Mechanical Systems and Signal Processing 87 (2017), pp. 294–306.

[111] Yu Wang, Yizhen Peng, Yanyang Zi, Xiaohang Jin, and Kwok-Leung Tsui. "A two-stage data-driven-based prognostic approach for bearing degradation problem". In: IEEE Transactions on industrial informatics 12.3 (2016), pp. 924–932.

[112] Lu Wang, Li Zhang, and Xue-zhi Wang. "Reliability estimation and remaining useful lifetime prediction for bearing based on proportional hazard model". In: Journal of Central South University 22.12 (2015), pp. 4625–4633.

[113] Wojciech Zulawinski, Katarzyna Maraj-Zygmat, Hamid Shiri, Agnieszka Wylomanska, and Radoslaw Zimroz. "Framework for stochastic modelling of long-term non-homogeneous data with non-Gaussian characteristics for machine condition prognosis". In: Mechanical Systems and Signal Processing 184 (2023), p. 109677. ISSN: 0888-3270.

[114] Hai Qiu, Jay Lee, Jing Lin, and Gang Yu. "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics". In: Journal of sound and vibration 289.4-5 (2006), pp. 1066–1090.

[115] J Lee, H Qiu, G Yu, J Lin, et al. "Rexnord technical services: Bearing data set". In: Moffett Field, CA: IMS, Univ. Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames (2007).

[116] Abdenour Soualhi, Hubert Razik, Guy Clerc, and Dinh Dong Doan. "Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system". In: IEEE Transactions on Industrial Electronics 61.6 (2013), pp. 2864–2874.

[117] Jaouher Ben Ali, Brigitte Chebel-Morello, Lotfi Saidi, Simon Malinowski, and Farhat Fnaiech. "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network". In: Mechanical Systems and Signal Processing 56 (2015), pp. 150–172.

[118] Yuqing Zhou, Anil Kumar, Chander Parkash, Govind Vashishtha, Hesheng Tang, and Jiawei Xiang. "A novel entropy-based sparsity measure for prognosis of bearing defects and development of a sparsogram to select sensitive filtering band of an axial piston pump". In: Measurement 203 (2022), p. 111997.

[119] Anil Kumar, Chander Parkash, Govind Vashishtha, Hesheng Tang, Pradeep Kundu, and Jiawei Xiang. "State-space modeling and novel entropy-based health indicator for dynamic degradation monitoring of rolling element bearing". In: Reliability Engineering & System Safety 221 (2022), p. 108356.

[120] Anil Kumar, CP Gandhi, Yuqing Zhou, Rajesh Kumar, and Jiawei Xiang. "Latest developments in gear defect diagnosis and prognosis: A review". In: Measurement 158 (2020), p. 107735.

[121] Qing Li, Steven Y Liang, Jianguo Yang, and Beizhi Li. "Long range dependence prognostics for bearing vibration intensity chaotic time series". In: Entropy 18.1 (2016), p. 23.

[122] Eric Bechhoefer and Rune Schlanbusch. "Generalized prognostics algorithm using Kalman smoother". In: IFAC-PapersOnLine 48.21 (2015), pp. 97–104.

[123] Eric Bechhoefer and Rune Schlanbusch. "Generalized prognostics algorithm using Kalman smoother". In: IFAC-PapersOnLine 48.21 (2015), pp. 97–104.

[124] C.K.R. Lim and D. Mba. "Switching Kalman filter for failure prognostic". In: Mechanical Systems and Signal Processing 52-53.1 (2015), 426 – 435.

[125] Lingli Cui, Xin Wang, Huaqing Wang, and Jianfeng Ma. "Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter". In: IEEE Transactions on Instrumentation and Measurement 69.6 (2019), pp. 2858–2867.

[126] J. F. Lawless. "Statistical models and methods for lifetime data". In: Canadian Journal of Statistics 10.4 (1982), pp. 316–317.

[127] Peter J Brockwell and Richard A Davis. Introduction to time series and forecasting. springer, 2016.

[128] Peter J. Rousseeuw and Christophe Croux. "Alternatives to the Median Absolute Deviation". In: Journal of the American Statistical Association 88.424 (1993), pp. 1273–1283.

[129] Yanyuan Ma and Marc G. Genton. "Highly Robust Estimation of the Autocovariance Function". In: Journal of Time Series Analysis 21.6 (2000), pp. 663–684.

[130] Christophe Croux and Peter J. Rousseeuw. "Time-Efficient Algorithms for Two Highly Robust Estimators of Scale". In: Computational Statistics. Ed. by Yadolah Dodge and Joe Whittaker. Heidelberg: Physica-Verlag HD, 1992, pp. 411–428. ISBN: 978-3-662-26811-7.

[131] Faicel Chamroukhi, Allou Samé, Gérard Govaert, and Patrice Aknin. "Time series modeling by a regression approach based on a latent process". In: Neural Networks 22.5-6 (2009), pp. 593–602.

[132] Faicel Chamroukhi, Allou Samé, Gérard Govaert, and Patrice Aknin. "A regression model with a hidden logistic process for feature extraction from time series". In: 2009 International Joint Conference on Neural Networks. IEEE. 2009, pp. 489–496.

[133] Peter J Huber. "Robust statistics". In: International encyclopedia of statistical science. Springer, 2011, pp. 1248–1251.

[134] James O Street, Raymond J Carroll, and David Ruppert. "A note on computing robust regression estimates via iteratively reweighted least squares". In: The American Statistician 42.2 (1988), pp. 152–154.

[135] S. Gross and W. L. Steiger. "Least absolute deviation estimates in autoregression with infinite variance". In: Journal of Applied Probability 16.1 (1979), 104–116.

[136] G Samorodnitsky and MS Taqqu. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman and Hall, 1994.

[137] Robert B Randall, Jérôme Antoni, and S Chobsaard. "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals". In: Mechanical Systems and Signal Processing 15.5 (2001), pp. 945–962.

[138] Paweł Zimroz, Hamid Shiri, and Jacek Wodecki. "Analysis of the vibro-acoustic data from test rig-comparison of acoustic and vibrational methods". In: IOP Conference Series. Earth and Environmental Science. Vol. 942. 1. IOP Publishing. 2021.

[139] Valeria CMN Leite, Jonas Guedes Borges da Silva, Giscard Francimeire Cintra Veloso, Luiz Eduardo Borges da Silva, Germano Lambert-Torres, Erik Leandro Bonaldi, and Levy Ely de Lacerda de Oliveira. "Detection of localized bearing faults in induction machines by spectral kurtosis and envelope analysis of stator current". In: IEEE Transactions on Industrial Electronics 62.3 (2014), pp. 1855–1865.

[140] Hamid Shiri and Jacek Wodecki. "Analysis of the sound signal to fault detection of bearings based on Variational Mode Decomposition". In: IOP Conference Series: Earth and Environmental Science. Vol. 942. 1. IOP Publishing. 2021, p. 012020.

[141] Reza Izanloo, Seyed Abolfazl Fakoorian, Hadi Sadoghi Yazdi, and Dan Simon. "Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise". In: 2016 Annual Conference on Information Science and Systems (CISS). IEEE. 2016, pp. 500–505.

[142] Badong Chen, Xi Liu, Haiquan Zhao, and Jose C Principe. "Maximum correntropy Kalman filter". In: Automatica 76 (2017), pp. 70–77.

[143] Gang Wang, Rui Xue, and Jinxin Wang. "A distributed maximum correntropy Kalman filter". In: Signal Processing 160 (2019), pp. 247–251.

[144] Guoqing Wang, Yonggang Zhang, and Xiaodong Wang. "Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems". In: Signal Processing 163 (2019), pp. 87–94.

[145] L.C.K. Reuben and D. Mba. "Diagnostics and prognostics using switching Kalman filters". In: Structural Health Monitoring 13.3 (2014), 296 – 306.

[146] M Kevin. "Learning switching Kalman filter models, 98-10". In: Compaq Cambridge Research Lab Tech Report (1998).

[147] NasraAbdalla Ahmed Mustafa et al. "Design of Smart Wearable System for Sudden Infant Death Syndrome Monitoring". PhD thesis. Sudan University of Science and Technology, 2021.

[148] Lim Chi Keong Reuben and David Mba. "Diagnostics and prognostics using switching Kalman filters". In: Structural Health Monitoring 13.3 (2014), pp. 296–306.

[149] Wojciech Żuławiński, Katarzyna Maraj-Zygmąt, Hamid Shiri, Agnieszka Wyłomańska, and Radosław Zimroz. "Framework for stochastic modelling of long-term non-homogeneous data with non-Gaussian characteristics for machine condition prognosis". In: Mechanical Systems and Signal Processing 184 (2023), p. 109677.

[150] P Zarchan and H Musoff. "Polynomial Kalman filters". In: Fundamentals of Kalman Filtering: A Practical Approach (2005), p. 156.

[151] Mikel Arrinda, Mikel Oyarbide, Haritz Macicior, Eñaut Muxika, Hartmut Popp, Marcus Jahn, Boschidar Ganev, and Iosu Cendoya. "Application dependent end-of-life threshold definition methodology for batteries in electric vehicles". In: Batteries 7.1 (2021), p. 12.

[152] Eric A Wan and Rudolph Van Der Merwe. "The unscented Kalman filter for nonlinear estimation". In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373). IEEE. 2000, pp. 153–158.

[153] Maria Isabel Ribeiro. "Kalman and extended kalman filters: Concept, derivation and properties". In: Institute for Systems and Robotics 43.46 (2004), pp. 3736–3741.

[154] Gotz Trenkler. "Statistical distributions: M. Evans, N. Hastings and B. Peacock (1993): (2nd Edition) ". In: Computational Statistics & Data Analysis 19.4 (1995), pp. 483–484.

[155] John P. Nolan. "Numerical calculation of stable densities and distribution functions". In: Communications in Statistics. Stochastic Models 13.4 (1997), pp. 759–774.

[156] Aleksander Weron and Rafal Weron. Computer simulation of Lévy α-stable variables and processes. Springer Berlin Heidelberg, 1995, pp. 379–392.

[157] B. L. Welch. "'Student' and Small Sample Theory". In: Journal of the American Statistical Association 53.284 (1958), pp. 777–788. ISSN: 01621459.

[158] Student. "The Probable Error of a Mean". In: Biometrika 6.1 (1908), pp. 1–25.